Evolutionary history of the Lesser Egyptian Jerboa, *Jaculus jaculus*, in Northern Africa using a multi-locus approach

Ana Filipa da Silva Moutinho

Mestrado em Biodiversidade, Genética e Evolução

Departamento de Biologia

2015

Orientador

Zbyszek Boratyński, Post-Doc Researcher, CIBIO-InBIO

Coorientadores

Paulo Célio Alves, Associate Professor, FCUP/ CIBIO-InBIO

Joana Paupério, Post-Doc Researcher, CIBIO-InBIO

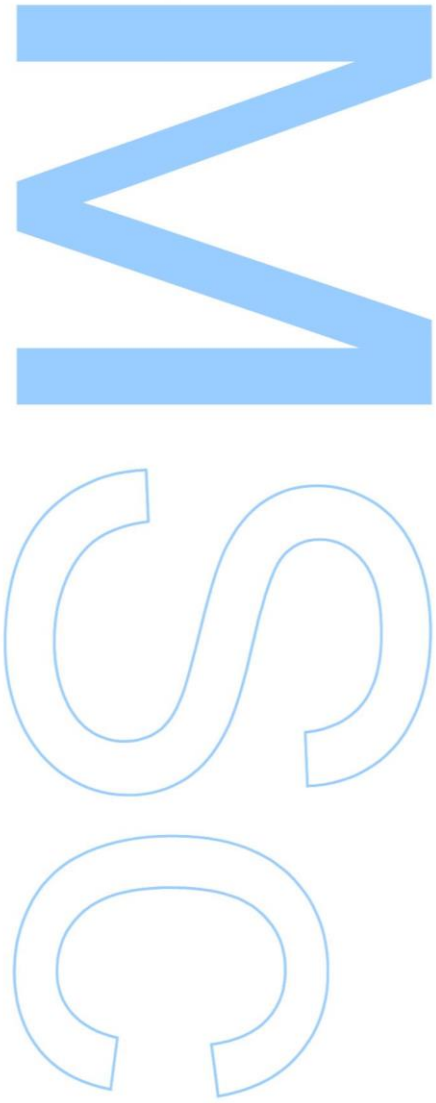
U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



Agradecimentos

First, I would like to thank my coordinator, Dr. Zbyszek Boratyński, for all the support along the work and opportunities that he made possible. I want also thank him for the amazing experience in Morocco during our field expedition.

Ao Professor Doutor Paulo Célio Alves, por ter possibilitado este projeto e pelo apoio fornecido.

À Doutora Joana Paupério pelo apoio extraordinário durante todo o processo, foi essencial!

Ao Doutor José Carlos Brito pelo apoio prestado desde o início.

Aos membros do Biodeserts, por todo o apoio e conselhos fornecidos. Um especial obrigado ao Paulo pela ajuda numa das análises, e à Teresa por toda a paciência e apoio essencial ao longo destes anos todos.

À Susana Lopes, por toda a ajuda e tempo disponibilizado no processo de desenvolvimento e otimização dos microssatélites.

À Clara, por ter efectuado uma das leituras independentes na análise dos microssatélites.

Aquele obrigado especial à Sara e ao Fábio, vocês tornam tudo muito mais fácil. Os três mosqueteiros estarão sempre juntos!

Aos meus colegas de mestrado, a vossa amizade e apoio foram indispensáveis.

A todos os outros membros do CTM e CIBIO que me ajudaram no laboratório ou noutros contextos.

To Kristian and Alison for the amazing time in Morocco.

À minha família, por estar sempre presente, pelo apoio imprescindível durante todo este percurso, e pela incrível paciência nos momentos mais complicados. Sem vocês eu não estaria onde estou hoje! E claro, à pantufa, por toda a companhia.

A todos os meus amigos, em especial à Carolina, à Maria e à Kati, por todos os momentos incríveis e por sempre acreditarem em mim.

Sumário

As alterações climáticas e os seus efeitos na biodiversidade são uns dos tópicos de discussão eminente entre a comunidade científica. As frequentes variações no clima têm consequências significativas nas fronteiras das regiões do Saara e do Sahel, levando a modificações periódicas na composição dos habitats, o que, por sua vez, influencia a distribuição, ecologia e evolução da fauna endémica das regiões desérticas. Esta dinâmica ao nível climático pode induzir variações em pressões seletivas e potenciar o isolamento filogeográfico das populações, conduzindo a eventos de diversificação genética e de adaptação, podendo assim ativar processos de especiação. Neste contexto, espécies hábeis em ambientes desérticos, como os roedores da família Dipodidae que habitam o Norte de África, adquiriram a atenção de investigadores devido à sua ampla distribuição geográfica, desde o Saara Atlântico até à Península Arábica, e ao seu elevado polimorfismo genético e fenotípico. Estudos anteriores constataram a existência de duas linhagens evolutivas divergentes entre a espécie reconhecida como *Jaculus jaculus*, "The Lesser Egyptian Jerboa", no Noroeste Africano, com distribuições geográficas extensas e sobrepostas. Foi proposto que as duas linhagens mitocondriais poderiam estar em isolamento reprodutivo, formando assim duas espécies crípticas estreitamente relacionadas. No entanto, o nível de fluxo genético entre estas duas linhagens nunca foi avaliado. Deste modo, este estudo tem como objetivo averiguar o nível de diferenciação entre as linhagens de *Jaculus* ao nível das diferentes regiões do genoma, estudar a sua história evolutiva e estimar os processos que conduziram à especiação, dando ênfase ao nível de isolamento reprodutivo entre as duas linhagens. Através de uma análise com base em múltiplos *loci* independentes, usando genética populacional, métodos filogenéticos e filogeográficos, os resultados confirmam a elevada divergência entre as duas linhagens de *J. Jaculus*, bem como a existência de um reduzido fluxo génico entre estas, indicando fortes sinais de isolamento reprodutivo. Não obstante, os níveis significativos de fluxo de genes observados sugerem que este processo possa estar incompleto. De qualquer forma o nível de divergência detetado é comparável ao normalmente encontrado entre espécies, sendo o tempo de separação estimado coincidente com as grandes mudanças climáticas nas regiões do Norte de África durante a transição do Plioceno para o Pleistoceno. A diversidade genética detetada em cada linhagem sugere histórias demográficas distintas, sendo que a Linhagem 1 apresenta um momento de expansão mais antigo que a Linhagem 2. Além disso, o efetivo populacional mais elevado observado para a Linhagem 1 está de acordo com a

potencial vantagem competitiva proposta para esta linhagem em comparação com a Linhagem 2. De acordo com as normas taxonómicas, as duas espécies crípticas podem ser identificadas como *J. jaculus* e *J. deserti*, como tem sido proposto na literatura. Este estudo sugere que o processo de especiação ocorreu na presença de fluxo génico, e que, possivelmente, a adaptação local teve um papel fulcral na incrível diversidade genética observada entre as espécies. Contudo, são necessários estudos adicionais para se obter uma melhor avaliação dos possíveis mecanismos geográficos e ecológicos subjacentes à evolução e especiação nos jerboas.

Palavras-chave: espécies crípticas, demografia, fluxo génico, adaptação local, seleção natural, filogenética, genética populacional, isolamento reprodutivo, especiação.

Abstract

Climate change and its effects on biodiversity is currently one of the most prominent scientific topics. The frequent shifts in climate caused significant changes in the Sahara-Sahel boundaries, leading to periodic modifications of the composition of habitats, thus influencing the distribution, ecology and evolution of the desert biota. Such dynamics resulted in new selective pressures and/or phylogeographic isolation of populations, causing events of genetic diversification, adaptation and eventually speciation. In this context, desert specialists such as African jerboas acquired the attention of researchers due to their broad geographic distribution across the Saharan-Arabian extent, and their great phenotypic and genetic polymorphism. Previous studies have recognized the existence of two divergent mitochondrial lineages within the Lesser Egyptian Jerboa (*Jaculus jaculus*), with a sympatric distribution in North West Africa. It has been proposed that the lineages could be reproductively isolated, forming two closely-related cryptic species. However, it was never evaluated if the reproductive isolation between clades indeed evolved, and if so, to what extent it had halted gene flow. In this respect, the present study aimed to verify the genome-wide signal of the differentiation between lineages, assess the evolutionary history of the species and estimate the potential processes driving speciation, thus highlighting the level of reproductive isolation between the putative species. By applying a comprehensive approach based on multiple independent loci and using population genetics, phylogenetic and phylogeographic methods, we validated the occurrence of two sympatric cryptic species within the Lesser Egyptian Jerboa. Both Structure and *Isolation-with-Migration* analyses showed very low frequency of gene flow between lineages, therefore suggesting strong, although not complete, reproductive isolation between species. The divergence level estimated is comparable to the one generally found between species, with the splitting age coinciding with the major climate shifts in North African regions during the Late Pliocene/Early Pleistocene boundary. The intra-lineages genetic variation suggested divergent demographic histories, where Lineage 1 experienced expansion slightly earlier than Lineage 2. Furthermore, the higher effective population size detected for Lineage 1 suggests an enhanced performance in occupying wider ranges when compared to Lineage 2, probably influenced by its competitive advantage over the micro-habitat preferences. According to the taxonomic norms, the two putative species could be named as *J. jaculus* and *J. deserti*, as has been previously proposed. We suggest that the process of speciation occurred in the presence of gene flow, wherein local adaptation probably had the key role in enhancing

the recovered genetic diversity between species. However additional studies should be performed to further evaluate the possible geographic and ecological mechanisms behind the evolution and speciation in jerboas.

Keywords: cryptic species, demography, gene flow, local adaptation, natural selection, phylogenetics, population genetics, reproductive isolation, speciation.

Index

Figure Index

Table index

Abbreviations

Mya: Million years ago;
GPS: Global Position system;
DNA - Deoxyribonucleic acid;
ºC: Celsius degrees;
µL: microliter;
ng: nanogram;
PCR: Polymerase Chain Reaction;
v: version;
indels: insertion/deletions polymorphism;
ML: Maximum-Likelihood;
BI: Bayesian Inference;
EBSP: Extended Bayesian Skyline;
MCMC: Markov chain Monte Carlo;
mtDNA: mitochondrial DNA;
Myr: Million years;
bp: base pairs;
TD: Touchdown procedure;
LD: Linkage disequilibrium;
HWE: Hardy-Weinberg Equilibrium;
BP: before present;
MRCA: Most Recent Common Ancestor;
Mid: Middle;
IM: isolation-with-migration model;
sp.: species;
F_{IS}: Inbreeding coefficient;
F_{ST}: Fixation index of genetic differentiation;
F_{IT}: Total Fixation Index;
He: Expected heterozygosity;
Ho: Observed heterozygosity;
uHe: unbiased expected heterozygosity;
SE: Standard error;
Na: Number of Alleles;
Ne: Number of effective alleles;
km – Kilometer;

min – Minute;

s: seconds;

P: probability;

N_e: Effective population size;

$2Nm_2$: The rate at which genes of the Lineage 2 are supplanted by genes from Lineage 1;

$2Nm_1$: The rate at which genes of the Lineage 1 are supplanted by genes from Lineage 2;

PCA: Principal Coordinate Analysis;

Cytb: cytochrome *b*;

DBX5: intron 5 from the developing brain, homeo box gene;

ADRA2B: alpha-2B adrenergic receptor;

GHR: growth hormone receptor;

IRBP: interstitial retinoid binding protein;

υWF: υon Willebrand factor;

Dxy: nucleotide divergence;

Da: net nucleotide divergence.

1. Introduction

1.1 Global climate changes and its role in shaping intraspecific genetic structure

Climate change and its effects on biodiversity are the most conspicuous scientific topics at the moment. Biodiversity can be defined as "the variation at all levels of biological organization" (Gaston & Spicer 2004) wherein the intraspecific genetic variation represents its most fundamental level, since it provides the basis for evolutionary changes (May 1994). Given the current global climatic changes, it is particularly essential to study the effects of climate alterations on the intraspecific genetic diversity if we intend to fully comprehend the evolutionary consequences of the ongoing climatic fluctuations and its long-term effects on biodiversity (Pauls *et al.* 2013).

During the past three million years the global climate has suffered great fluctuations. Major glaciations of the Quaternary period were underlined by the massive extension of polar ice sheets causing a huge impact on temperature, vegetation zones and mountain blocks, producing long land bridges in many regions of the world (Figure 1). In general, during glaciation periods species inhabiting mountains stepped down to lower altitudes, tropical rainforests were reduced and the areas covered with deserts and savannahs were extended (Colinvaux 1997). It is now clear that the effects of the major climatic shifts were different across the globe, inducing several regional differences in land cover and ocean currents. Likewise, species' responses varied according to their natural range, being always linked to local geographic and climatic conditions (Hewitt 2000).

Pleistocene is recognized as the epoch of the major climate oscillations, with a crucial impact on the patterns of population genetic variation within several species (Hewitt 1996). Yet, the way these cycles contributed to species diversification is still unclear. This uncertainty is justified by the lack of precision when estimating times of species divergence during such dynamic periods; not only due to the absence of sufficient genetic resolution to estimate differentiation (i.e. when analysis rely on a single genetic locus), but also associated with the analytical difficulties in acquiring accurate divergence time estimates when speciation occurred in the presence of gene flow (Carstens & Knowles 2007).

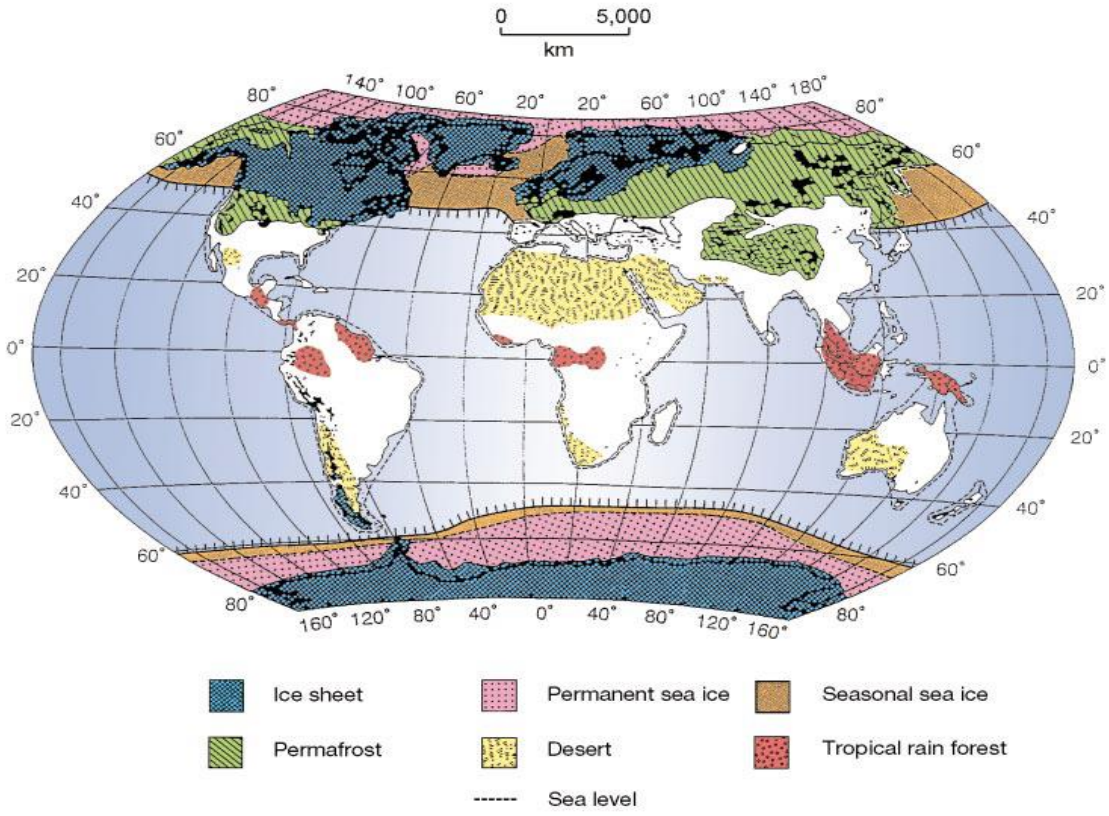


Figure 1. Map representing desert and tropical rain forest regions, lowered sea level and the extent of ice and permafrost at 20 000 years BP (Hewitt 2000).

Natural populations react to environmental change through ecological plasticity or adaptation (Riddle *et al.* 2008; Scoble & Lowe 2010; Hoffmann & Sgrò 2011), or by shifting their habitat or geographical distribution, hence modifying communities' composition and the nature of species relationships (Parmesan *et al.* 2003; Hoffmann & Sgrò 2011). Although organisms have the capacity to respond to such climate shifts, the extreme speed and magnitude of climate change has the potential to lead the most vulnerable populations, or even entire species, to extinction (Hoffmann & Sgrò 2011). However, whatever the response of a single population, or a species, to climate fluctuations, it will always affect the amount and distribution of intraspecific genetic diversity, most likely connected to a loss of genetic variation (Pauls *et al.* 2013).

Similarly to genetic variation, the adaptation potential varies across species' populations and its distribution ranges (Davis & Shaw 2001; Eckert *et al.* 2008). Therefore it is likely that single populations within a species will react in distinct ways to climate change (Procaccini *et al.* 2007), mainly if there are differences in the selective pressures across species' distribution (Hoffmann & Sgrò 2011). Invasive species, with the ability to rapidly adapt to new selective pressures outside their native ranges, are a

good example of diversified adaptation (Sax *et al.* 2007). Moreover, two recent studies focusing on distinct geographical types of *Arabidopsis thaliana*, adapted to local climate conditions, revealed that fitness is connected to specific climate-adapted genetic loci (Fournier-Level *et al.* 2011; Hancock *et al.* 2011). Such diversification in adaptation processes as a response to the same or different selective pressures might even promote the evolution of reproductive isolation (Schluter 2001; Via 2012; Pauls *et al.* 2013).

Diversified adaptation is also exemplified with the development of cryptic diversity in some populations and species (Rice 1987). This is particularly frequent in populations with wide and coincident distribution ranges, which may be exposed to distinct drivers of natural selection and thus subdivided into evolutionary discrete forms. Due to the absence of geographical barriers, these organisms might present few differences in morphology, being therefore cryptic. The identification and delimitation of these species complexes contributes for a better understanding of speciation as well as its underlying processes (Sites & Marshall 2003), being also fundamental to conservation and biodiversity management (Moritz 1994).

1.2. Species delimitation and the underlying processes of speciation

The adaptation of populations to contrasting environments is the primary process for the formation of new species (Bennett 1872; Darwin 1968; Schluter 2001). An extensive debate regarding species concepts and speciation has been going on for years since “The Origin of Species” (Darwin 1968). Indeed, defining species and acknowledging the processes behind their differentiation are of great importance for studies in biodiversity and evolutionary biology (Seifert 2009; Jowers *et al.* 2014). However, determining the processes underlying speciation and the traits that define a species is complex and controversial. For example, populations can constitute distinct ecological units (ecological species) and still continue to exchange genes, which does not establish an evolutionary species; while others can fully represent reproductive isolated units, thus evolutionary species, without the existence of ecological divergence between them, not being considered ecological forms (Schilthuizen 2000; Butlin *et al.* 2012).

The controversy around the delimitation of species is mainly driven by the large number of species concepts (Mayden 1997; de Queiroz 2005; Hey 2006; De Queiroz 2007) and the different backgrounds, or motivations, for research on speciation (Bird *et*

al. 2012). The majority of studies of sexually reproductive organisms continue to rely on the predominantly accepted "Biological Species Concept", defining species in terms of interbreeding, where reproductive incompatibility determines a species boundary (Wright 1940; Mayr 1942, 1963). From a different perspective, the "Evolutionary Species Concept" (Simpson 1951; Wiley 1978) defines a species as a set of unique evolutionary units, with different tendencies and historical fate. Mayden (1997, 1999) documented the latest as a lineage-based concept, setting aside all other definitions by labelling them as "secondary operational tools for species recognition" (Sites & Marshall 2003). De Queiroz (2005, 2007) composed adjacent arguments for a "general lineage concept" considering all the properties of diverging lineages, like reproductive isolation and ecological preferences, as determinants in species delimitation. He suggested a "unified species concept" where all the mentioned properties were rejected as essential and treated as terms of their significance as evidences to deduce species limits (Sites & Marshall 2003). Naomi (2011) accounts for an integrated framework of species delineation, combining the theoretical "Evolutionary Species Concept" (Wiley 1978; Mayden 1997, 1999) and the necessary components for delimiting species (intrinsic reproductive isolation, ecological preferences, reciprocal monophyly, etc; De Queiroz 2005, 2007). The concepts proposed by Mayden (1997, 1999), de Queiroz (2005, 2007) and Naomi (2011) are thought to be the most comprehensive, since they combine ideas both from species conceptualization and species delimitation, connecting genetics and ecology. The use of this concept offers a wider application to speciation studies (Bird *et al.* 2012), with significant importance for polymorphic species that lack evident boundaries, or distinctiveness, like cryptic species.

The most traditional way to categorize speciation processes is to use the geographical context, dividing it into allopatric, parapatric and sympatric. Allopatric speciation relies on the existence of an extrinsic barrier during divergence, causing the absence of connectivity between populations. Since Mayr (1942) highlighted the geographical context as the main driver underlying the formation of new species, allopatric speciation has been considered as the "null hypothesis", since it was seen as the most common and plausible mode (Coyne & Orr 2004). This process is usually classified depending on how the ancestral population was divided [i.e. through vicariance - separation, or extinction, of populations with continuous distributions through geographic events, such as mountain uplift or river flow (Rosen 1978); or by peripatric differentiation - a particular variation of allopatry, defined by two geographically distinct units: a founder population (smaller) and a source population (larger) (Bird *et al.* 2012)]; and if the consequent reproductive isolation was complete or

in secondary contact (alloparapatric speciation; Coyne & Orr 2004). Parapatric speciation denotes for a partial extrinsic barrier, so there is only negligible contact between diverging populations (Butlin *et al.* 2012). Sympatric speciation means that although no extrinsic barrier is present, a genetic barrier emerges within a population or a species. Examples of drivers of such differentiation are distinct ecological preferences and differences in the timing and habitat selection for mating (Butlin *et al.* 2008).

However this geographic interpretation of speciation forms is far from capturing the full complexity of spatial relationships that may occur among diverging populations (Butlin *et al.* 2008, 2012). Several authors (Templeton 1981; Kirkpatrick & Ravigné 2002) have proposed new methods to categorize the processes behind speciation, in which they considered two additional factors: the forces driving reproductive isolation and the genetic basis for isolation (Butlin *et al.* 2008). It is true that speciation is a long process that might take over many generations (up to millions; Coyne & Orr 2004) and so conditions may change along the route. In this way, Butlin (2008) proposed a process that comprises the interaction between the three components of the speciation process: a spatial context, which defines the extrinsic isolation; the evolutionary forces, such as mutation drift, natural and sexual selection, that may act within the geographical context and thus generate intrinsic isolation; and the genetic signature, which will vary depending on the individual species (Kirkpatrick & Ravigné 2002). Consequently, the genetic divergence between demes will be proportional to the age and the level of isolation (Jowers *et al.* 2014), and may result in reproductive isolation if sufficient genetic diversity is gathered (Rundle & Nosil 2005; Jowers *et al.* 2014).

This integrated method highlighted the idea of accepting sympatry as a geographical context anticipating speciation, and that reproductive isolation might develop even though there is a high probability of recurrent gene flow among populations (Bird *et al.* 2012). Indeed, this would be the most comprehensive way to go if we want to fully understand the effects of the evolutionary forces and the genetic architecture in shaping species boundaries, and the subsequent development of reproductive isolation. In this context, advances in genetics, in particular the development of multi-locus methods, have provided some of the most reliable methodologies when considering species delimitation and biodiversity assessments. By sampling genetic variability at different loci we obtain more accurate estimates of intraspecific genetic diversity, thus giving a straight forward evaluation of biodiversity, and allowing the identification of cryptic forms and biodiversity hotspots (Gómez & Lunt 2007).

1.2.1. Genetic methods – a comprehensive approach

Acquiring genetic data at the boundary between populations and species is extremely valuable in terms of species delimitation, particularly when analysing gene trees from neutral loci (Harrison 1998; Templeton 2001). In a given gene tree, the rate at which an ancestral polymorphism is lost enables valuable estimations of temporal divergence between sister lineages and the branching pattern of a phylogeny (Carstens & Dewey 2010). For single-locus data, gene trees may allow qualitative assessments of lineage limits, resulting from genealogies estimations (Pellegrino *et al.* 2005). However, single-locus approaches may not display the real patterns of lineage splitting and divergence, mainly due to the stochasticity nature of the coalescent theory (Rosenberg & Nordborg 2002). Instead, data from multiple loci have the ability to mitigate the random effects of the genetic drift, as new techniques to estimate the phylogenetic relationships between recently diverged lineages have emerged in recent years (Carstens & Dewey 2010).

Although there is still a high proportion of studies applying single-locus genetic methods, mainly relying on uniparental inherited markers (i.e. mitochondrial DNA), for phylogeographic and phylogenetic analysis (Avise *et al.* 1987; Moore & Aug 1995; Zhang & Hewitt 2003; Weiss & Ferrand 2007), it is becoming more clear that such datasets provide limited information, following a rough, or even entirely misleading, assessment of population history (Pamilo & Nei 1988; Wu 1991; Palumbi & Baker 1994; Hare & Avise 1998; Sequeira *et al.* 2008; Degnan & Rosenberg 2009; Boratyński *et al.* 2014b). Therefore, reliable inferences of population structure analyses, as well as population history estimations, should focus on multiple independent loci (Pritchard *et al.* 2000; Rosenberg & Nordborg 2002; Sequeira *et al.* 2008; Boratyński *et al.* 2014b).

Evidences show that analyses consisting of multiple independent markers with different evolutionary rates, such as mitochondrial, nuclear exons and microsatellites, afford a greater robustness regarding population history inferences, once they encompass information from different temporal scales (Estoup *et al.* 1995; Palo *et al.* 2004; Sequeira *et al.* 2008). Mitochondrial DNA (mtDNA) has been largely used in phylogenetic and phylogeographic analyses due to its relatively simple application, provided by the lack of recombination, assumed neutrality and smaller effective sample size, owing to its maternal inheritance (Zhang & Hewitt 2003; Hickerson *et al.* 2010). Compared to nuclear exons, mtDNA has an estimated mutation rate of five to ten times faster, yet still offering access to substantially distant time scales, which ranges up to million years (Wan *et al.* 2004). In turn, microsatellites, also known as short tandem

repeats (STR), present extraordinary evolutionary rates, that makes them extremely useful for inferences of recent phylogeographic events at the intraspecific level, such as modern genetic patterns of populations within species, as well as its genetic structure (Schlötterer 2004; Wan *et al.* 2004).

Regarding phylogenetic analyses, the individual gene trees and the underlying species tree might be substantially different. Several newly developed phylogenetic methods, incorporating coalescent models, have been looking for accurate estimations of species tree based on multiple loci datasets. These methodologies assume that incomplete lineage sorting, that is "the failure of two or more lineages in a population to coalesce, leading to the possibility that at least one of the lineages first coalesces with a lineage from a less closely related population" (Degnan & Rosenberg 2009), is the focal source of discordance between gene trees and species trees. It presumes no recombination within loci, free recombination between loci and no gene flow succeeding speciation (Degnan & Rosenberg 2009; Carstens & Dewey 2010; Heled & Drummond 2010). Methods of species tree estimation allow researchers to model the probability of individual samples' membership to the respective evolutionary lineage, enabling the direct inference of the relationships between sister lineages (Knowles & Carstens 2007; Carstens & Dewey 2010). Furthermore, by combining demographic analyses and assessments of gene flow estimations between closely related species we get an almost full comprehension of population structure and evolutionary history of the species under study.

The application of such comprehensive methods enables a higher robustness when estimating the past history of populations (Crandall *et al.* 2000). Moreover, these methodologies are particularly helpful to investigate relations between closely related species, giving light in the sense of cryptic speciation (Paupério *et al.* 2012).

1.3. Biodiversity dynamics in North Africa and Sahara-Sahel

North African regions are of great interest in terms of biogeographic efforts, thanks to its wide diversity of habitats, heterogeneous landscapes and complex paleoclimatic and geological patterns (Le Houérou 1997; Sayre *et al.* 2013). The great climate shifts that occurred during the Pliocene-Pleistocene interval (5.3 million years) were important mediators of the remarkable evolutionary changes of fauna and flora in the region. Evidences from the marine sediment sequences reveal that the subtropical African climate oscillated between periods of drier and wetter conditions, regulated by the orbital fluctuations of the Earth. These periods were marked by the increase of

step-like landscapes accompanied with a higher aridity, which were coincident with the onset and intensification of high latitude glacial cycles (deMenocal 2004).

North Africa comprises two of the main ecoregions of the continent: the Sahara desert and the neighbouring arid Sahel, located in the sub-Saharan regions. The Sahara is the largest warm desert in the world that alongside with the Sahel region covers around 11,230,000 km^2 (Olson *et al.* 2001), revealing a patchy structure with significant differences in topographical and climatic aspects. The boundaries along the Sahara and the Sahel establishes the shift between the Palearctic and Afrotropical biogeographic realms (Olson *et al.* 2001), which induces a great latitudinal disparity in species distribution and the increment of local biodiversity (Dumont 1982; Foley *et al.* 2003). The Sahara-Sahel spreads over ten countries, in which many of them are labelled as low development (UNDP 2010) under a long-term political instability, that hampers the accomplishment of field surveys, trans-border research and conservation planning (Brito *et al.* 2014).

Deserts are characterized by their aridity; however they have huge differences in abiotic characteristics. The variability between deserts is probably higher than between any other biomes, greatly because deserts are widely spread along the planet and have arisen from very distinct reasons (Ward 2009). These regions are commonly classified as homogenous with low diversity patterns, drawing less attention for scientific research (Durant *et al.* 2012). However, recent studies using novel sampling and genetic methods started to increase the knowledge on species composition and distributions in these habitats, uncovering cryptic diversity and, in some cases, lineage splitting of what was once considered as wide-ranging species. Such studies are allowing an entirely different view regarding regional biodiversity patterns (Brito *et al.* 2014).

Desert conditions of the Sahara are believed to have started around 7 million years ago (Mya) in Chad (Schuster *et al.* 2006), or around 6 to 2.5 Mya in western areas (Swezey 2009). Ever since the desert first appeared in North Africa, during the Mid-Upper Pliocene, these regions have been through intense fluctuations in space and time, shifting between more humid and more arid phases (Rognon 1993; Foley *et al.* 2003; Kröpelin *et al.* 2008; Drake *et al.* 2011), in response to climatic oscillations triggered by modifications in the Earth's orbital processes, sea surface temperature and responses mechanisms of vegetation cover due to the variation in rainfall (Wang *et al.* 2008; Claussen 2009) (Figure 2). It is believed that the Sahara-Sahel regions went through eight to ten wet and dry periods over the past 125,000 years (Le Houérou 1997). These fluctuations in climate and land-cover have caused significant changes in the Sahara-Sahel boundaries, leading to periodic modifications of desert biota (Dumont

1982; Brito *et al.* 2014) and alterations in the ecological compositions of landscapes. Such dynamics resulted in new selective pressures and/or phylogeographic isolation, causing events of genetic diversification, adaptation and eventually speciation (Mouline *et al.* 2008; Brito *et al.* 2014).

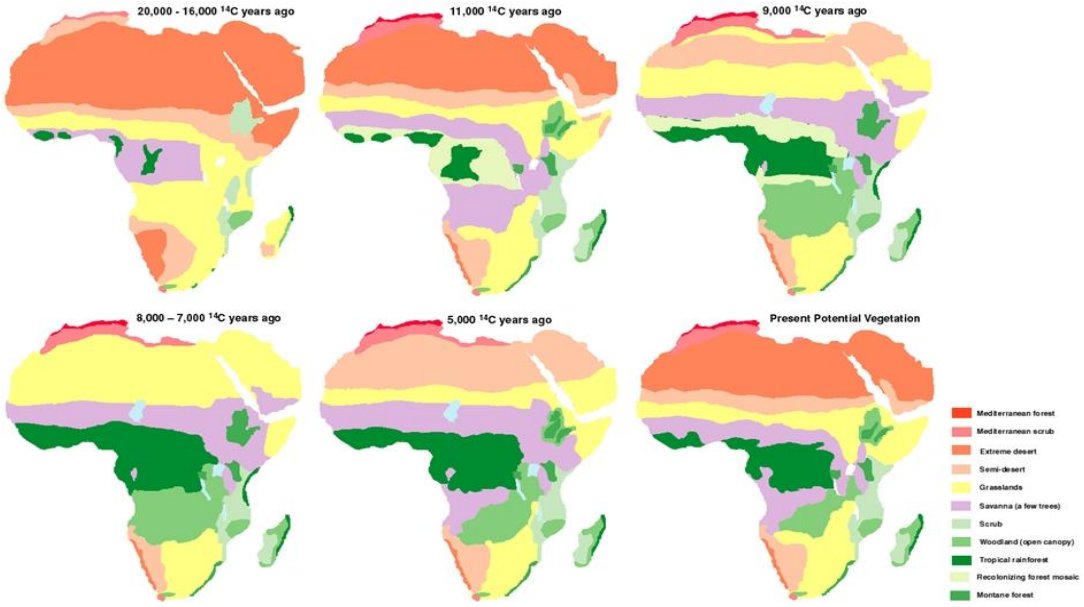


Figure 2 . Temporal and spatial fluctuations in African habitats since the Last Glacial Maximum (Adams & Faure 2004).

1.3.1. The evolutionary processes in Sahara-Sahel

Despite the few phylogeographic studies done within the region, there are some hypotheses presenting the Sahara-Sahel as a promoter of taxa diversification since the onset of the arid conditions and throughout the successive range shifts (Douady *et al.* 2003; Carranza *et al.* 2008). Considering a neutral scenario (without adaptation processes) as the principal driver of speciation, species divergence emerged from vicariant events, where the following allopatric effects induced the interruption of gene flow, leading to the development of evolutionary independent lineages or even new species. The time and nature of these vicariant events have diversified effects on the different taxa, in accordance with their habitat requirements. For example, xeric species might have experienced diversification processes during humid periods (e.g. *Jaculus* spp.; Boratyński *et al.* 2012), while multiple mesic organisms (adapted to arid environments but still needing some humidity) went through population contraction and

diversification along with hyper-arid conditions (e.g. *Agama* spp.; Gonçalves *et al.* 2012) (Brito *et al.* 2014).

Population expansion and contraction during the shifts between favourable and harsh climate conditions led to the identification of three distinct patterns: 1) dispersal through the spatial corridors that favoured gene flow during suitable climate periods; 2) divergence without gene flow in the geographical refuges, promoting speciation; and 3) the evolutionary change, and adaptation as a response to harsh climatic conditions. Mountains are a fundamental character when discussing diversification processes across Sahara-Sahel, since they act as refugia for several species and enable gene exchange under favourable climatic conditions. Hence, hard climatic conditions endorsed the diversification of multiple taxa in these regions, perhaps proceeding in long-term allopatric isolation and speciation (Brito *et al.* 2014).

At the same time, the extreme conditions that define the desert environment might be also responsible for modelling the evolution and diversification processes of different species (Boratyński *et al.* 2012). In this respect, it is proposed that due to these conditions in the deserts many organisms that inhabit the region developed extreme adaptations (i.e. physiological ability to survive without access to water) which allow them to occupy vast areas where such conditions occur (Haim & Izhaki 1995). However, these exceptional adaptive characteristics can be seen as a limitation for their distribution because it may prevent them to disperse across other than desert habitats (i.e. habitats too humid). This might cause subdivisions of populations during periods of climate change as revealed by the shifts between more humid and more arid phases in Sahara-Sahel regions (Brito *et al.* 2014).

1.3.2. African Jerboas

In the context of the biodiversity dynamics in Sahara-Sahel, jerboas have drawn attention of researchers because of their vast and broad distribution across the Saharan-Arabian extent and their high phenotypic and genetic polymorphism (Ben Faleh *et al.* 2010a; Boratyński *et al.* 2012). African Jerboas (*Jaculus* spp., Erxleben 1777, Dipodidae) are rodents adapted to extremely dry conditions, inhabiting variable types of deserts and semi-arid areas of south Palearctic and Afrotropical regions, from Senegal (*Jaculus jaculus*) in West Africa to Afghanistan and Pakistan (*Jaculus blanfordi*) in the East (Wilson & Reeder 2005). The genus, comprising three different species, *Jaculus jaculus*, *Jaculus blanfordi* and *Jaculus orientalis*, is recorded from the Late Miocene (Zhang *et al.* 2012) and through time substantial intraspecific variability

was observed, conducting to the acknowledgement of several subspecies or species (Shahin 2003; Wilson & Reeder 2005; Shenbrot *et al.* 2008). Due to its wide and heterogeneous distribution, this group is of great interest for studies on the relations among environmental change, geological events and speciation history (Zhang *et al.* 2012).

Jaculus jaculus (Linnaeus, 1758), commonly designated as Lesser Egyptian Jerboa, has a broad distribution in North Africa, occupying the Sahara-Sahel, the Horn of Africa and also the Middle East, comprising the Arabian Peninsula and Central Asia (Wilson & Reeder 2005; Amori *et al.* 2008; Aulagnier et al, 2009; Ben Faleh *et al.* 2012b) (Figure 3). It is characterized by a strictly nocturnal activity, feeding on seeds, insects and the moist parts of deserts grasses which it detects by using its acute sense of smell (Granjon & Duplantier 2009). Jerboas do not need to drink water to survive on the deserts, as they can fulfil their water needs from its diet. It is also featured by jump locomotion, specifically reflected in the hind limb morphology (Amori *et al.* 2008). It has the ability to travel great distances to look for food, up to 20 km per day (Granjon & Duplantier 2009), which it easily makes due to its large feet and hopping stride (Figure 4). It occupies diverse habitats through its range, from sandy dunes to rocky substrates. However they are always found close to vegetation. *J. jaculus* usually inhabits burrows, which during the end of spring and summer are well hidden and sealed with a plug of sand during the day, to keep the humidity inside, making these places ideal for the individuals to rest, avoid predators and escape from the heat of the day (Granjon & Duplantier 2009).

Numerous studies on morphology (Shahin 1999, 2005), karyology (Granjon *et al.* 1992; Ata & Shahin 1999, 2006; Ata *et al.* 2001; Shahin & Ata 2001, 2004) and biochemistry (Shahin 2003) were done in order to access its taxonomic status. Within this species the definition of subspecies was first based on morphological data (Heim De Balsac 1936). For North Africa, particularly in Egypt, four subspecies were described - *J. jaculus jaculus*, *J. jaculus botleri*, *J. jaculus flavillus*, and *J. jaculus schluteri* (Osborn & Helmy 1980); and a total of five in Libya (Ranck 1968) and Algeria (Heim De Balsac 1936; Corbet 1978) - *J. jaculus deserti*, *J. jaculus centralis*, *J. jaculus sefrius*, *J. jaculus airensis*, and *J. jaculus jaculus* (Ben Faleh *et al.* 2010b). However, in most cases, morphological studies were not followed by ecological and genetic surveys, and so no inferences could be made regarding potential adaptive signatures of variation and the evolutionary history of the lineages (Shahin 2003).

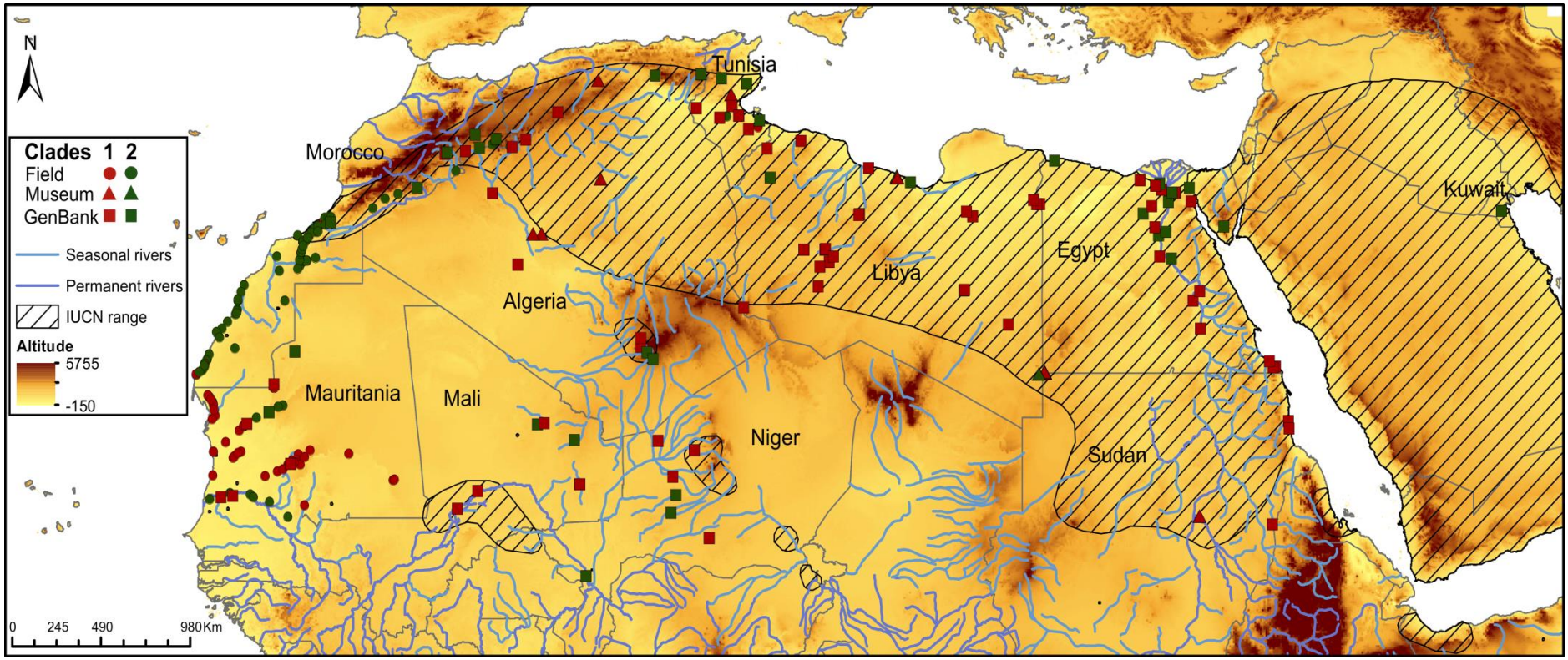


Figure 3. Geographic locations of the lesser Egyptian jerboa specimens used in the present study. Red and green samples denote, respectively, the first and the second lineage previously described. Field (circles) and museum samples (triangles) are specified in the map, as well as the GenBank sequences downloaded (square; see Annex 1 and 3 for detailed information). The known distribution range according to the IUCN is also shown (black line filled area; IUCN 2008). Each geographic region is labelled according to the respective country.

1.3.2.1. *Jaculus jaculus* species complex

Since the last 50 years, the Lesser Egyptian Jerboa has been receiving attention regarding the emergence of two cryptic species with a broad and sympatric distribution, corresponding to two different colour forms with putative divergent ecological preferences. In Algeria, Peter (1961) recorded two "ecological forms" distinguished by the coat colour: a dark grey form inhabiting rocky areas and a yellow-brownish type restricted to sandy areas. In Libya, Ranck (1968) documented two sympatric species, *Jaculus jaculus* (Linnaeus 1758) and *Jaculus deserti* (Loche 1867), matching two colour variants: *J. jaculus* with a pale orangish dorsum with whitish-grey hair roots; and *J. deserti* characterized by a dark dorsum with grey hair roots. He also reported other features like the colour of the hind foot sole hair (white versus dark), the number of foramina in the angular process (one versus two) and the size of the skull (large versus smaller) and of the auditory bulla (more or less inflated). However, Harrison (1978) claimed that the two forms were most likely conspecific populations, being improperly segregated by Ranck's assumptions. Relying on the same arguments, Corbet (1978) gave similar suppositions by including *J. deserti* as a subspecies of *J. jaculus*, which has been widely acknowledged among taxonomists (Wilson & Reeder 2005). Notwithstanding, the ongoing morphological studies continue to question the existence of only one species within the widespread *J. jaculus* in North Africa. Gharaibeh (1997) again proposed the conception of two different sympatric forms with distinct ecological preferences. He performed an overview of the morphological and ecological information obtained in previous surveys, but still, maintained the same taxonomic designation (*Jaculus jaculus*) for all lesser jerboas in Tunisia.

Recent studies have evidenced a close correlation among morphological and genetic variation (Ben Faleh *et al.* 2010a, b, 2012; Boratyński *et al.* 2012, 2014a). Ben Faleh and colleagues (2010a) used a mitochondrial DNA molecular marker (cytochrome *b*), morphometrics (measurements of the size of the skull) and frequencies of 23 loci encoding 16 enzymatic proteins to investigate the taxonomy of the species. They proposed the existence of two different lineages within the population of *Jaculus* sp. in Tunisia, which reasonably matched two morphometrically distinct groups. However, the study of enzymatic proteins was not successful in defining species-diagnostic loci, which could be explained by the presence of hybridization and gene flow, homoplasy or incomplete lineage sorting, or by the lack of variation in those markers (Ben Faleh *et al.* 2010a). Thus, it was unclear to define whether the differences amongst nuclear and mitochondrial markers were merely due to a higher polymorphism in mtDNA or whether nuclear genomes were being

homogenized (e.g. by gene exchange between species) (Boratyński *et al.* 2012). Moreover, this genetic and morphometric differentiation was not followed by karyological differences (Ben Faleh *et al.* 2010b).

Latter studies, focusing on mitochondrial and on one nuclear gene (Ben Faleh *et al.* 2012; Boratyński *et al*. 2012, 2014a) supported the existence of the two divergent lineages previously described in Tunisia, within *J. jaculus,* in North West Africa, with a broad and sympatric distribution, yet with an unclear, or absent, geographic structure. Although divergence in the nuclear marker was not that strong, it still differentiated the two lineages. Since no signs of introgression were detected, reproductive isolation was hypothesized between these two mitochondrial clades (Boratyński *et al.* 2012). Additionally, mitochondrial analysis revealed the emergence of a third clade potentially confined to the Middle East, which may be closely-related with one of the lineages (Ben Faleh *et al.* 2012b).

More recently, Boratyński and co-workers (2014a) conducted a study that supported the former assumptions, by detecting a linkage between a putatively adaptive trait (dorsal fur colour) and the colour of the local habitat, which might also lead to the observable phenotypic differences between the genetic lineages. Those results suggest that the sympatric clades might persist in ecological separation within the mixture of sandy (lighter) and rocky (darker) micro-habitats over North Africa, with one of the clades being linked to brighter and sandy areas, and the other to the darker rocky substrates (Figure 4). These outcomes suggest specific micro-habitat preferences that may have contributed to reproductive isolation, allowing the persistence of separate demes (Boratyński *et al.* 2014a).

Despite the evidences of a putatively adaptive trait related to the dorsal fur colour, the phenotypic variation found between lineages is rather continuous than dichotomous. Therefore, the admixture of habitat and geographic ranges is followed by a highly overlapping phenotypic variation between lineages, thus implying the evolution of cryptic diversity (Boratyński *et al.* 2014a).

Although several hypotheses have been proposed, linking this remarkable divergence to gene flow or reproductive isolation, which implies a speciation event; deep knowledge of the evolutionary processes behind this extraordinary variability within the Lesser Egyptian Jerboa is still unclear and controversial.

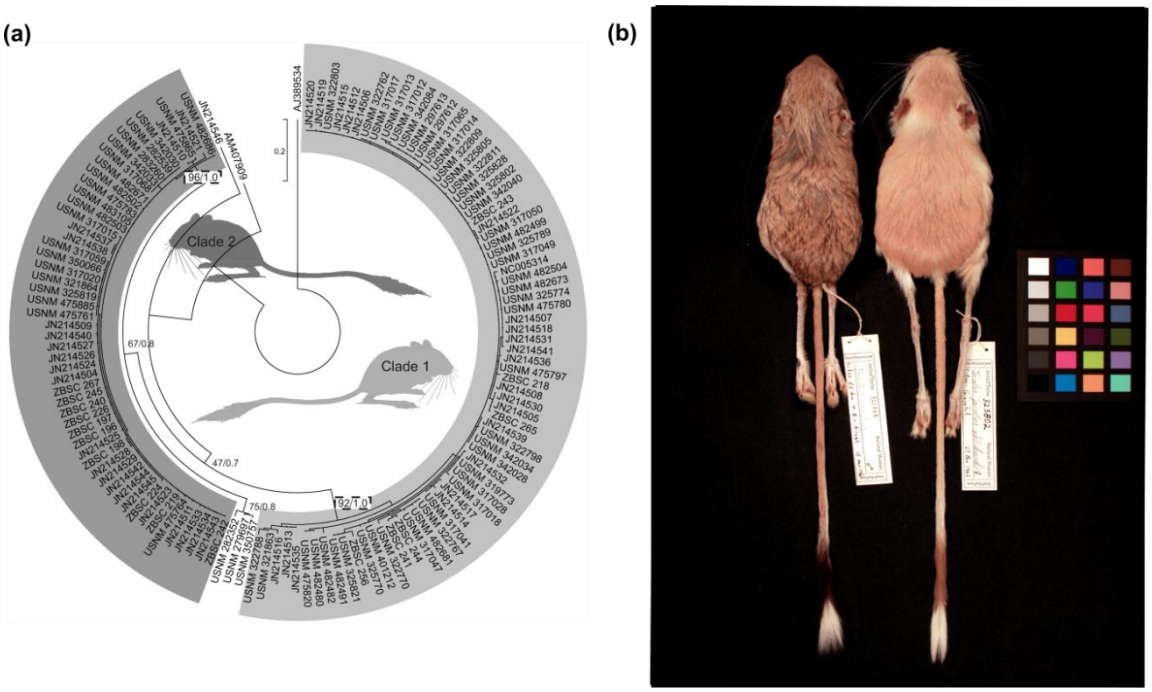


Figure 4. Genetic and phenotypic variation between the two lineages described. (**a)** Maximum likelihood (ML) tree of a fragment of the cytochrome b gene, displaying the bootstrap values and posterior probabilities (Bayesian methods) that support divergence between clades**; (b)** A photography representing the colour variation observed within African jerboas, defining two putative cryptic species (Boratyński *et al.* 2014a)**.**

1.4. Objectives

The present study intends to assess the evolutionary history of the Lesser Egyptian Jerboa and determine the level of reproductive isolation between the two putative cryptic species described, so that a better understanding of the evolutionary processes behind the extraordinary variability found between them can be achieved. To do so, we propose the use a multi-locus approach based on the development and analysis of multiple microsatellites markers and several nuclear and mitochondrial genes sequencing. The study encompasses all the North African range, focusing mainly on populations from Morocco, Western Sahara and Mauritania regions, where both clades occur mostly in geographic sympatry, though revealing some levels of parapatry. In this survey we focused on three main objectives:

1) Investigate the genome-wide phylogenetic signal of the differentiation of the two putative cryptic species, by analysing several independent markers with different

modes of inheritance and using species delimitation and species tree inference methods;

2) Assess the evolutionary history of *Jaculus* species, through the estimation of divergence time and demographic history of the two putative species;

3) Estimate current genetic structure and determine the levels of admixture of the two putative species through the analysis of microsatellite data and the application of isolation-with-migration (IM) models.

This will give the basis for the validation, or rejection, of the two separate species of jerboa, contributing to a better resolution of the phylogeographic pattern and speciation process of this species, until now recognized as the Lesser Egyptian Jerboa, *Jaculus jaculus*.

2. Material and Methods

2.1. Study area and field work

The study area comprises all the North African range, covering almost all the distribution extent of the species complex. However, the collection of tissue samples focused mainly on populations from Morocco, Western Sahara and Mauritania regions (field samples; Figure 3).

In 2015, we conducted an expedition to a remote area of Morocco's Atlantic Saharan region from February 13th to 26th (Moutinho *et al.* 2015). Sampling methods of jerboas included night capturing with spotlights and hand nets, and the identification of road carcasses (road killed animals). All individuals were photographed and their geographic locations recorded with a GPS (Nomad, Trimble) (Figure 5). Specimens were preserved in 96% ethanol and are stored at the Natural History Museum of the Département de Zoologie et Ecologie Animale, Institut Scientifique de Rabat, Morocco. All procedures were approved by le Haut Commissariat des Eaux et Forêts et de la Lutte Contre la Désertification (Direction de la Lutte Contre la Désertification et la Protection de la Nature, decision no: 42/2014).

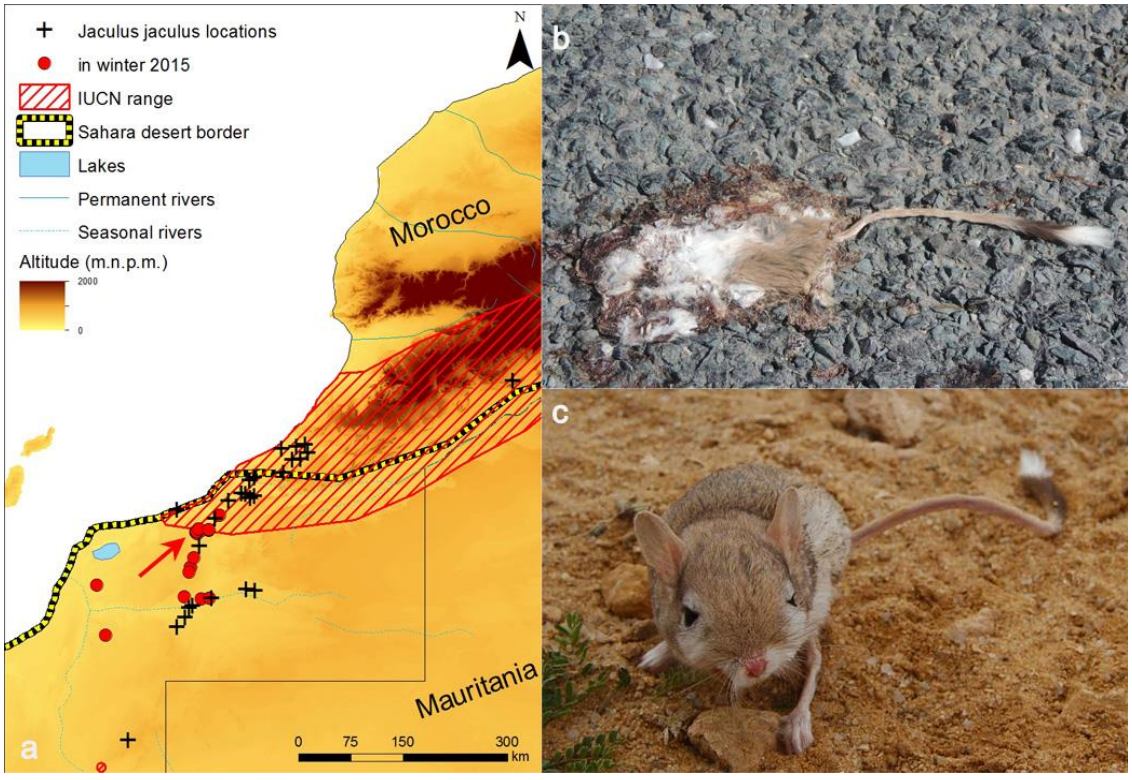


Figure 5. (a) Distribution map of *Jaculus jaculus* in the Moroccan Atlantic Sahara. Red circles represent detected locations of species occurrence during winter 2015 expedition. The red arrow indicates populations under intensive breeding during winter 2015. Black crosses represent species occurrence detected in previous expeditions. The

distribution range according to the IUCN is shown (red line filled area; IUCN 2008); **(b)** Photograph of a road killed specimen (by AF Moutinho); **(c)** Photograph of a caught in hand individual (by Z Boratyński; modified from: Moutinho *et al.* 2015)**.**

Throughout the expedition we captured a total of 14 individuals and found 11 road killed specimens of *J. jaculus* (Figure 5). This field survey allowed the identification of the highest population density in an area with relatively high amounts of vegetation, a possible hotspot for breeding events since all the captured females from this area were reproducing. These findings increased the limited knowledge of species reproductive biology (Aulagnier *et al.* 2009; Boratyński *et al.* 2012), highlighting the potential for population growth during winter seasons (Moutinho *et al.* 2015).

2.2. Sampling and DNA extraction

A set of 231 samples distributed throughout North Africa, comprising 152 tissue samples from field surveys and 79 samples obtained from various Museum Collections around Europe and USA, were used in this study (see Annex 1).The tissue samples were collected from road-killed and live animals during several overland field expeditions in North-West Africa or offered by other researchers and collaborators (see Annex 1 for detailed information). Specimens were sampled from November 2011 to February 2015, corresponding to the trip conducted in Morocco this winter (Moutinho *et al.* 2015). Tissue samples were preserved in 96% ethanol for genetic analyses at the moment of collection. The Geographical position of each sample was recorded with a Global Positioning System (GPS) on the WGS84 data and represented using the Geographical Information System ArcMap 10.0 (Figure 3). Of these 152 samples, 54 samples were already used in previous studies, but were analysed for other markers (Boratyński *et al.* 2012, 2014a; Annex 1). Additionally 10 samples of *J. orientalis* were used in analysis as outgroups (Annex 1).

Extractions of the genomic DNA from tissue samples collected in the expeditions were performed using EasySpin Kit, following the “Genomic DNA Minipreps Tissue Kit” protocol. Extractions of museum samples were performed in a separate and autonomous facility, under sterile conditions, using the QIAamp® DNA Micro Kit (QIAGEN), following the “Isolation of Total DNA from Nail Clippings and Hair” protocol. Extracted DNA was stored at -20ºC.

2.3. Mitochondrial and Nuclear DNA sequences analyses

2.3.1. Amplification and sequencing

The genetic diversity of *J. jaculus* was first assessed at the mitochondrial level using Cytochrome *b* (*Cytb*) for the entire set of samples, giving continuity to previous studies assessing genetic variability in this species complex (Ben Faleh *et al.* 2010, 2012a, 2012b; Boratyński *et al.* 2012, 2014a). A set of 51 samples had already been used in previous studies (Boratyński *et al.* 2012, 2014a; see Annex 1), but were re-sequenced in order to obtain a larger fragment. Amplification of the partial *cytb* gene (897 bp) was accomplished using two primer pairs previously designed for *Jaculus* species (Jac1Fw, Jac1Rv, Jac4Fw, Jac4Rv; Boratyński *et al.* 2012). Museum samples imply fragmented DNA, thus hindering the amplification method (Taberlet & Luikart 1999; Taberlet *et al.* 1999). Therefore, the reconstruction of the DNA fragment with the museum samples was done in steps to produce overlapping sequences in order to produce the entire fragment. In some cases, only a short fragment (325 bp) of the gene was amplified, which was obtained combining two pairs of primers (Jack4Fw and Jack1Rv) (Primers, references and PCR conditions in Annex 2).

Nuclear loci were amplified for a smaller set of samples, containing only specimens collected in the field trips (152 samples). Amplifications involved several optimization procedures. Of the sexual chromosomes genes a total of 5 markers were tested: intron 5 and 6 from the *DBX* gene on the X-chromosome and introns *DBY7*, *UBE1Y* and *SMCY7* on the Y-chromosome (Hellborg & Ellegren 2003, 2004). After several amplifications tests only intron 5 from the *DBX* gene was successfully optimized. Additionally 4 autosomal genes were optimized and amplified: exon 10 from *GHR* (*growth hormone receptor*), exon 1 from *ADRA2B* (*alpha-2B adrenergic receptor*), exon 1 from *IRBP* (*interstitial retinoid binding protein*) and exon 28 from *υWF* (υ*on Willebrand factor*). The locus *υWF* was previously amplified for 21 samples (Boratyński *et al.* 2012; see Annex 1), although these samples were re-sequenced to obtain the fragment size analysed in the present study. Thereby, 5 nuclear loci were used in analyses (Primers and PCR conditions in Annex 2).

Amplifications were performed in a final volume of 10 µl using 5 µl of QIAGEN PCR MasterMix, 0.4 µl of each primer and 1 µl of DNA (approximately 10 ng of genomic DNA), varying according to DNA quality. PCRs were carried out in a thermocycler MyCycler (BIO-RAD). A negative control was used in all PCRs to check for contaminations. The successful PCR products were purified by applying ExoSAP (USB® ExoSAP-IT® PCR Product Cleanup, Affymetrix), as specified by the manufacturer, to remove the primers and nucleotides that were not used in

amplification process (reaction of 30 min: 15 min at 37ºC and 15 at 85ºC). After this, sequencing reactions were performed following the standard protocol of Big-Dye cycle sequencing kit (BigDye® Terminator v3.1 Cycle Sequencing Kits, AB Applied Biosystems). Sequences were obtained for both strands on an ABI 3130xl Genetic Analyser (AB Applied Biosystems). For some of the genes, sequencing of both strands was performed in an external laboratory (Macrogen Inc.).

2.3.2. Sequence alignment and phylogenetic analyses

Sequences were verified and aligned using SEQSCAPE v2.6 (Kosman *et al.* 2001). Available sequence data for the *cytb* gene, of our target species (150 sequences) was downloaded from GenBank and included in the analyses (see Annex 3). Alignments for each locus were generated with CLUSTAL W (Thompson *et al.* 1994) implemented in ClustalX v2.0 (Larkin *et al.* 2007) and edited manually in BIOEDIT v7.1.3 (Hall 1999) so that an increase of blocks of base pairs identity could be achieved while clustering insertion/deletions (indels). Polymorphic positions for each of the discrete sequences from nuclear loci were carefully examined to ensure precise and consistent identification of double peaks in heterozygotes. Moreover, heterozygous positions for insertion/deletions mutations were resolved manually from offset chromatogram peaks (Flot *et al.* 2006). Haplotypes for the *cytb* gene were inferred with DnaSP v5 (Librado & Rozas 2009). For further analysis, haplotypes for each nuclear locus were inferred using PHASE v2.1 (Stephens *et al.* 2001; Stephens & Scheet 2005) with three runs performed for each locus with 10,000 burn-in steps and 10,000 interactions. Input files were created in SEQPHASE (Flot 2010). Individual sequences holding insertions/deletions were phased manually and included in SEQPHASE as *known haplotype pairs*. PHASE results were consistent throughout runs for all loci. Haplotypes presenting probability phase calls below 80% were discarded from the analysis to ensure reliable haplotype estimations. The insertions/deletions observed in the X intron were coded manually (see Annex 4) due to the large size of these polymorphisms (21/42 base pairs) and were used in network reconstructions but were excluded from other analyses.

Phylogenetic analysis was first performed individually for each locus analysed. The Akaike information criterion (AIC) was used to select the best-fit model of sequence evolution for each locus alignment among the 88 available in the software jModelTest v2.1.4 (Darriba *et al.* 2012) (see Annex 5 for detailed information of the specific evolutionary model for each locus). The phylogenetic relationships between

haplotypes were inferred by the Maximum-Likelihood (ML) approach in PHYML v3.0 (Guindon *et al.* 2010) and the Bayesian phylogenetic inference (BI) implemented in MrBayes v3.2.0 (Ronquist *et al.* 2012). ML analyses were performed with 1000 bootstrap pseudoreplicates. Bayesian posterior probabilities were assessed from two runs with four chains of 1 million generations for the nuclear genes and two runs with five chains of 50 million generations for cytochrome *b*, with a sampling frequency that provided a total of 10,000 samples for each run, discarding 25% of burn-in. Tracer v1.5 (Drummond & Rambaut 2007) was applied to check the ESS values (effective sample size) for each analysis. Resulting trees were drawn with FIGTREE v1.3.1 (Rambaut 2009).

Phylogenetic networks for each gene individually were generated using parsimony calculations in TCS v1.21 (Clement *et al.* 2000) considering gaps as a fifth state. Each indel of the *DBX5* locus was considered as a single mutational step, regardless of the corresponding size (see Annex 4). Analyses were performed for each locus with a connection limit of 95%, which means that below this threshold haplotypes are not connected. *Cytb* and *DBX* genes presented haplotypes disconnected and so networks were redrawn with the connection limit fixed at 90 steps in order to link the more unrelated groups and see the number of mutational steps among them. Networks were edited manually and in tcsBU (Santos *et al.* 2015). Additionally, neighbour-net networks based on uncorrected patristic distances and bootstrap analysis with 1000 replicates were done for *cytb* and for the combined 5 nuclear locus (3740 bp) in SPLITSTREE v4.13.1 (Huson & Bryant 2006).

2.3.3. Species tree inference and molecular dating

Gene trees may be considerably different from the fundamental species tree. The phylogenetic methods used to determine the species tree assume that lineage sorting is the main cause of discordance between gene trees and species trees and that there is no recombination within loci, free recombination between loci and absence of gene flow succeeding speciation (Carstens & Dewey 2010; Heled & Drummond 2010). Following the assumptions of these methods, alignments were first tested for the presence of recombinant locus in SPLITSTREE where three of the genes (*DBX5*, *υWF* and *IRBP*) evidenced significant probability for recombination. These three loci were then analysed with IMgc (Woerner *et al.* 2007) to reduce the dataset to the largest non-recombinant blocks.

The species tree was inferred by applying the multispecies coalescent model that is implemented in *BEAST (Heled & Drummond 2010), part of the BEAST v2.3.0 package (Bouckaert *et al.* 2014). The same outgroup sequences were used for this analysis. The input file was produced with the application BEAUti v2.3.0, part of the BEAST package. Individuals were assigned to the corresponding taxa according to the two mitochondrial lineages previously described (Ben Faleh *et al.* 2010a, b, 2012b; Boratyński *et al.* 2012, 2014a). Preliminary analyses were carried out to evaluate the clock-like evolution of each locus by inspecting the posterior distribution of the standard deviation of an uncorrelated lognormal relaxed clock. Based on these trial runs the final analysis was accomplished with the HKY+I+G substitution model for *cytb* under an uncorrelated lognormal relaxed clock. Analyses of the nuclear loci were carried out with a HKY (+I for *υWF* and *ADRA2B*) substitution model under a strict molecular clock (see Annex 5 for detailed information).

Times of divergence were estimated using *cytb* as the reference gene. A fossil-based calibration rate for genetic evolution was not possible due to the poor fossil record of *Jaculus* in North Africa. Similarly, the well-known calibration point Muridae-Rodentia was not used due to the likely saturation effect associated with the ancientness of the divergence between Muridae and Dipodidae. Instead, a universal value for *cytb* evolutionary rate was applied after an extensive rodent literature review (Triant & DeWoody 2006; Nabholz *et al.* 2008; Ben Faleh *et al.* 2012a). The clock rate was fixed at 0.02 per millions of years with a standard deviation of 0.005. Considering a normal distribution, these settings are in agreement with a central 95% range of about 0.010–0.030 substitutions per Million years (Myr), considering one year of generation time.

Following these assumptions the prior of the relaxed clock standard deviation was set to a normal distribution with a mean of 0.02 with *sigma* fixed at 0.005. The coalescent constant population size was used as tree prior and all the remaining priors were set to defaults. Three independent runs of 300 million generations were implemented, sampling trees and parameter estimators every 30,000 generations for all loci. The convergence of the runs was verified after the appropriate removal of a 10% burn-in using TRACER v1.5 (Drummond & Rambaut 2007). Visual inspection of trace plots indicated a good sample of all parameters for each *BEAST independent runs, with effective population sizes (ESS) above 1000, indicating a good convergence of all parameters. The results from all runs were combined with LogCombiner v2.3.0, and the subsequent maximum clade credibility summary trees with posterior probabilities for each node were generated with TreeAnnotater v2.3.0 from the BEAST

package. All the mentioned trees were visualized and edited with FIGTREE v1.3.1 (Rambaut 2009).

2.3.4. Population genetics and demographic analyses

Total and net divergences between lineages were calculated using Kimura 2-parameter distances (Kimura 1980) in MEGA v5.10 (Tamura *et al.* 2011). Additionally, the divergence among several rodent species, based on published data (Haynes *et al.* 2003; Jaarola *et al.* 2004; Montgelard *et al.* 2008; Blanga-Kanfi *et al.* 2009; Bannikova *et al.* 2009; Edrey *et al.* 2012; Lebedev *et al.* 2012; Paupério *et al.* 2012; Pisano *et al.* 2015), was inferred for comparison analysis. The same rodent species were not available for all loci, so different comparisons were performed. Standard deviations for these divergences were estimated from 10,000 bootstrap replications. Nucleotide diversity (π), θ_W (computed from the number of segregating sites) and haplotype diversity were calculated per lineage and for the whole population for each locus. Three neutrality test statistics, Tajima's *D* (Tajima 1989), Fu's F_S (Fu 1997) and R_2 (Ramos-Onsins & Rozas 2002) were calculated to detect recent population expansion. Significance was evaluated through 10,000 coalescent simulations. These statistics were assessed per locus for each lineage and for the whole *Jaculus* population with DnaSP v5 (Librado & Rozas 2009). Calculations were made separately for the entire data set and for the non-recombinant portions obtained with IMgc.

The dynamics of effective population sizes through time of the recognized populations of *J. jaculus* was inferred through Extended Bayesian Skyline Plots (EBSP; Heled & Drummond 2008), using a linear model in BEAST v2.3.0 (Bouckaert *et al.* 2014). EBSP uses Markov Chain Monte Carlo (MCMC) genealogy sampling to estimate the posterior distribution of effective population size over time under a multi-locus approach by using the times of coalescent events between gene trees (Heled & Drummond 2008). The input file for the analyses was obtained with the application BEAUti v2.3.0. The same non-recombinant dataset used for species tree inference was analysed. The evolutionary models for each locus of each lineage were estimated in jModelTest v2.1.4 (Darriba *et al.* 2012), which resulted in similar models to the previously obtained for the full dataset (see details in annex 5). The same mutation rate as above (0.02 nucleotide substitutions per million years; Triant & DeWoody 2006; Nabholz *et al.* 2008; Ben Faleh *et al.* 2012a) was applied to the mitochondrial DNA (*cytb*) under a strict molecular clock. The evolutionary rates of the nuclear loci were estimated according to the reference gene, *cytb*, under a strict molecular clock. The

prior for the mean distribution of population sizes was optimized according to the population sizes estimated in preliminary runs with a coalescent prior and a constant population size (Heled & Drummond 2008). Remaining priors were set as defaults. The MCMC parameters were the same applied in *BEAST analysis. TRACER v1.5 (Drummond & Rambaut 2007) was used to assess the convergence of the independent runs. Trace plots indicated a good sample of all parameters as the ESS for the independent and for the combined runs were higher than 200. Results of the separate runs were combined with LogCombiner v2.3.0, part of the BEAST package, after discarding 10% burn-in.

2.3.5. *Isolation-with-Migration* Analyses

Models of isolation-with-migration (IM) (Nielsen & Wakeley 2001) implemented in the IMa2 software (Hey 2010a) were applied to infer whether gene flow has occurred between the two described sympatric populations. This method estimates the multi-locus effective population sizes (for present and ancestral populations), divergence times and migration rates under a model of isolation with migration (IM) (Nielsen & Wakeley 2001; Hey & Nielsen 2004). This model assumes no recombination within each locus, free recombination among all loci, no population structure within each species or putative species, no genetic contribution coming from unsampled populations and that the sampled genes are under selective neutrality, not necessarily meaning strict neutrality (Hey 2010a; Pinho & Hey 2010). Following the assumptions, the non-recombinant dataset was used for the estimations. For each pair of populations being compared in analysis, IMa estimates values of $\theta=4N\mu$ (wherein μ is the mutation rate per year for each marker) for populations 1 and 2, and their ancestor, as well as the time (t) (where $t=t\mu$, the time of populations splitting at t generations in the past) that they diverged in the presence of gene flow (where $m=m/\mu$, given in the coalescent approach).

Analyses were structured considering the two lineages of *J. jaculus* as populations. Our main aim with these analyses was to assess whether the apparent low phylogenetic signal in nuclear markers is better enlightened by incomplete lineage sorting as a result of recent population splitting times ($m=0$) or by the presence of gene flow ($m>0$). To corroborate the consistency of the estimates across different runs, 2 independent analyses were performed, wherein the 6 loci were set with different starting seeds and parameters of upper bound priors. Locus specific mutation rates were estimated from the average number of substitutions per site (Dxy) and the net

nucleotide divergence (Da) among *J. jaculus* and *J. orientalis* specimens, calculated using Kimura 2-parameter distances in MEGA v5.10, considering a split time of 5.97 Myr (Pisano *et al.* 2015) and a generation time of 1 year (Nabholz *et al.* 2008). The geometric mean of the locus-specific mutation rates was used to calculate the effective population sizes and divergence times from the IMa2 highest posterior density of each parameter, following instructions in the manual (Hey 2010b). IMfig (Hey 2010a) was used to produce a graph were the estimates of the evolutionary history are represented as a phylogenetic tree made with boxes, symbolizing the sampled and ancestral populations, splitting times and migration.

2.4. Microsatellite analyses

2.4.1. Microsatellite selection and optimization

A microsatellite library for *J. jaculus* species was developed with high-throughput genomic sequencing (454 pyrosequencing) at GenoScreen (http://www.genoscreen.fr/en/). No microsatellites were published before for this species or closely-related species. Microsatellite retrieving and primer design was based on individuals from distinct geographic regions from North Africa. However, primer design was performed taking into account mainly one of the described mitochondrial lineages, thus complicating the amplification process due to the great variability observed between them. The aim was to design and optimize microsatellite markers that would amplify on both lineages of our target species. A total set of 40 loci was initially chosen from the database considering the quality rate (primers classified as “best” were favoured), repeat motif (balanced motifs were preferred), high number of repeats (above 11), small estimated product size [sizes below 200 bp due to the limitations of museum samples under PCR amplification methods (Taberlet *et al.* 1999); although these samples were not used in this study, it is one of the aims for future research] and melting temperatures. The selected loci were arranged in multiplexes for optimization. Multiplexes were designed combining markers with similar melting temperatures (maximum range of 5°C) of forward and reverse primers, and with a minimal interval of around 50 to 70 bp of the product size between loci labelled with the same dye colour (FAM, VIC, NED and PET) in order to avoid overlapping of the loci. AUTODIMER v1.0 (Vallone & Butler 2004) was used to perform screening analyses for intramolecular hairpins structures and primer dimer formation within each multiplex to avoid high complementary between primers, thus promoting a higher amplification success.

An exhaustive process of optimization of microsatellite markers was developed. Initially seven samples, representative of the two mitochondrial lineages described, as well as the distribution range of this study, were selected for the first steps of the optimization process. These selected samples revealed a high amplification success both on mitochondrial and nuclear genes, thus allowing a robust evaluation of the adjustments done to primer concentrations and to amplifications protocols. PCR reactions were accomplished using a Touchdown method in order to cover the range of the annealing temperatures.

Uniplex PCRs were performed when markers failed to amplify in multiplex reactions in order to estimate the loci individual amplification success, optimal conditions and genotypes profile. This method excluded several low-quality markers in the initial phase of the optimization process.

A total of 40 markers were tested. We started by testing two multiplexes reactions of 10 markers each. During the process more than a half of the candidate loci were discarded due to their week amplification and unreadable genotype profile. New markers were selected to increase the number of loci. They were combined with the previous multiplexes and further testing procedures were carried out. Some of the markers were shifted between the two multiplexes or had their label changed to improve efficiency. Such shifts allowed the maintenance of functional loci and beneficiated performance.

After optimization we ended up with two multiplexes, one with 7 and the other with 4 markers, and 2 loci that had to be amplified individually in separate PCR reactions (see Annex 6).

2.4.2. Genotyping

A total of 148 samples were genotyped for 13 microsatellite loci. Multiplex and individual reactions, primers concentrations and amplification conditions are summarized in Annex 6. All PCRs were performed in 10µl reaction volumes. Multiplex reactions consisted of 5µl of QIAGEN PCR Master Mix, 3µl of pure water, 1µl of primer mix and 1µl of template DNA. Individual PCRs included 5µl of QIAGEN PCR Master Mix, 2.8µl of pure water, 0.4µl of each primer (forward and reverse), 0.4µl of fluorescent tail and 1µl of template DNA. A negative control was used in every reaction. All PCRs were performed using a touchdown (TD) protocol, which varied on each multiplex and individual reaction conditions (see Annex 6). Each PCR reaction was carried out with an initial denaturation step at 95ºC of 15 min, a denaturation at 95ºC for 30s, a TD

method (-0.5ºC) with the respective annealing temperature (Annex 6) of 60s, an extension at 72ºC of 30 s, followed by a final extension of 30 min at 60ºC. The number of cycles for each multiplex/individual reaction is displayed in Annex 6. PCR products were run on an ABI 3130xl Genetic Analyser (AB Applied Biosystems) using 1µl of the amplified product for 10µl of formamide +75-400 (-250) LIZ NEW size standard.

Allele data was obtained using GENEMAPPER v4.0 (Applied Biosystems 2006). Sizing bin windows were manually created by comparison to the allelic ladder. Automated scoring was checked by three independent observers in order to minimize genotyping errors. Inconsistent genotypes were considered as missing data.

2.4.3. Data analysis

The final dataset consisted of 132 samples, 42 from lineage 1 and 90 from lineage 2 (details in Annex 1), as all samples displaying at least 40% of missing data were excluded from analysis. MICROCHECKER v2.2.3 (Van Oosterhout *et al.* 2004) was used to assess the presence of genotyping errors due to null alleles and allele dropout. Linkage disequilibrium (LD) and deviations from Hardy-Weinberg Equilibrium (HWE) were estimated with GENEPOP on the Web (http://genepop.curtin.edu.au/) and GENALEX v6.0 (Peakall & Smouse 2006). Confidence intervals for the analysis were inferred with the Bonferroni correction [0.05/(number of populations*number of loci)]. The sampling was continuous across all the distribution of the species and there were no geographical populations delimitated. Therefore, the analysis was performed considering only the two main lineages as populations. Six loci (Jac04, Jac07, Jac11, Jac12, Jac24, and Jac27) revealed significant deviations to Hardy-Weinberg Equilibrium and to Linkage disequilibrium assumptions. These results were expected considering Hardy-Weinberg expectations, since the two lineages do not reflect real populations. This is explained by one of the assumptions of Hardy-Weinberg Equilibrium which assumes random mating and the non-existence of genetic migration within a population (Guo & Thompson 1992). Analyses also revealed linkage between some of the loci which means that they are segregating together and so we cannot consider the information reliable. Therefore these six loci were discarded from further analysis. Moreover, one of the loci (Jac01) only amplified with samples belonging to Lineage 1, and so it was discarded due the large amount of missing data.

Measures of genetic diversity and variation, such as allele frequencies, mean number of alleles sampled per locus and population, observed (Ho) and expected (He) heterozygosity, allelic richness per locus and population and measures of Fst (θ) were

estimated with FSTAT v1.2 (Goudet 1995). F-statistics and individual-by-individual genetic distances that were then used to compute a Principle Coordinate Analyses (PCA) were calculated with GENALEX v6.0 (Peakall & Smouse 2006).

The number of clusters or populations and the level of putative admixture between lineages were inferred with the Bayesian Clustering software STRUCTURE v2.3.3 (Pritchard *et al.* 2000). It does not consider the original populations designation of each individual and clusters them inside a defined set of populations (K) that may or may not corroborate the initial assumptions. Analyses were accomplished by applying the admixture model with correlated allele frequencies. The software was run for a number of clusters (K) between 1 and 10 with 5 repetitions of 1,000,000 MCMC iterations for each K value, following a burn-in period of 100,000 steps. Three independent analyses were performed to ensure that similar posterior probabilities of the data in each run were obtained and to establish high levels of confidence in the model fit. STRUCTURE HARVESTER v0.6.92 online software (Earl & vonHoldt 2012) was used to determine the probability of each K value. The most likely number of clusters (populations) was assessed using the mean values of likelihood [L(K)] and the Evanno method (Delta K; Evanno *et al.* 2005).

3. Results

3.1. Phylogenetic analyses

In mitochondrial analyses (*cytb*), from a total of 152 field samples a set of 140 samples were successfully amplified for the short fragment (325 bp) and 137 samples for the long fragment (897 bp). Out of the 79 museum samples, 35 short fragments were amplified, but only 4 of the long fragment (see Annex 1). Thus, analyses with the short fragment were achieved with a total of 175 samples plus 150 sequences downloaded from GenBank, whereas the ones with the long fragment were undertaken with a set of 139 specimens plus 71 downloaded sequences from GenBank (Table 1; and see Annex 3 for detailed information about GenBank sequences). Concerning the outgroup specimens, all samples successfully amplified with the short fragment of *cytb*, while the long fragment was amplified on 7 samples (Annex 1).

The phylogenies based on both the short and long fragments of the mtDNA (*cytb*) retrieved two strongly supported and well defined monophyletic clades (Figure 6 and see Annex 7 for the short fragment phylogenetic tree inference and Annex 8 for the phylogenetic network), corroborating the two previously identified mitochondrial lineages (Ben Faleh *et al.* 2012; Boratyński *et al.* 2012, 2014a). Since both phylogenies generated similar topologies, further analyses were conducted only with the long fragment of cytochrome *b*. The analysis of the long fragment was accomplished with 156 haplotypes comprising 222 polymorphic sites denoting two highly diversified clades (see Table 1). All sequences were most likely of mitochondrial origin and not of nuclear integrated copies since they perfectly aligned with the complete mtDNA sequence of *J. jaculus* (GenBank accession number: NC_005314), and no insertion/deletions polymorphisms and or stop codons were found within the dataset.

Nuclear loci were only amplified for the field samples, so a total of 152 samples was amplified. *ADRA2B* and *GHR* fragments were effectively amplified for 137 samples; sequences of *IRBP* and *υWF* exons were obtained for 130 and 132 samples respectively; and the intron 5 from the *DBX* gene was efficiently amplified for 135 specimens (see Annex 1). All analysis of nuclear loci were performed with 6 successfully amplified outgroup sequences, being the same for all genes (see Annex 1). The number of sequences used for nuclear loci after the removal of the low probability phase calls, and the respective number of haplotypes and polymorphic positions are summarized in Table 1.

Table 1. Levels of polymorphism estimated at each locus in the whole population of *J. jaculus,* and for each lineage. For loci where significant levels of recombination were detected, polymorphism values are shown both for the recombinant and non-recombinant datasets (denoted as "without recomb'").

Locus	Lineage	L	N	S	H	Hd(SD)	π (SD)%	θ_W (SD)%	D	F_S	R_2
Cytb	***J. jaculus***	892	210	222	156	0.992(0.002)	5.20(0.11)	4.20(0.90)	0.38	-68.61***	0.10
	Lineage 1	894	87	148	79	0.997(0.003)	1.60(0.09)	3.20(0.90)	-1.84*	-77.91	0.04***
	Lineage 2	895	123	100	77	0.978(0.006)	0.50(0.03)	2.00(0.50)	-2.43**	-34.48	0.02***
DBX5	***J. jaculus***	305	264	18	12	0.554(0.026)	1.30(0.06)	0.90(0.30)	1.08	3.35	0.12
	Lineage 1	311	84	3	4	0.220(0.059)	0.09(0.03)	0.20(0.10)	-1.02	-1.96	0.04
	Lineage 2	306	180	7	8	0.208(0.040)	0.07(0.01)	0.40(0.20)	-1.77*	-8.94***	0.02
***DBX5* without recomb**	***J. jaculus***	295	255	17	10	0.541(0.026)	1.42(0.06)	0.90(0.30)	1.28	4.98	0.12
	Lineage 1	301	84	2	3	0.217(0.057)	0.07(0.02)	0.10(0.10)	-0.71	-0.95	0.06
	Lineage 2	296	171	6	7	0.166(0.038)	0.06(0.01)	0.40(0.20)	-1.73	-8.21***	0.02
ADRA2B	***J. jaculus***	693	252	15	19	0.641(0.032)	0.20(0.02)	0.40(0.10)	-0.86	-7.23	0.06
	Lineage 1	693	72	7	9	0.705(0.031)	0.30(0.02)	0.20(0.20)	0.32	-0.87	0.13
	Lineage 2	693	180	11	11	0.345(0.045)	0.06(0.01)	0.30(0.10)	-1.85*	-9.67*	0.02
GHR	***J. jaculus***	798	268	20	20	0.530(0.029)	0.09(0.007)	0.40(0.10)	-2.05*	-20.51*	0.02*
	Lineage 1	798	88	10	11	0.378(0.070)	0.05(0.01)	0.20(0.10)	-2.09*	-12.09**	0.03*
	Lineage 2	798	180	11	10	0.147(0.036)	0.03(0.008)	0.20(0.09)	-2.12*	-12.93***	0.02*

IRBP	***J. jaculus***	1058	156	40	54	0.967(0.005)	0.55(0.02)	0.70(0.18)	0.52	-32.75	0.07
	Lineage 1	1058	50	25	19	0.905(0.023)	0.35(0.05)	0.53(0.20)	-1.07	-6.48	0.07*
	Lineage 2	1058	106	23	35	0.948(0.010)	0.39(0.002)	0.41(0.13)	-0.19	-19.37	0.09
***IRBP* without recomb**	***J. jaculus***	681	133	23	22	0.877(0.016)	0.43(0.02)	0.62(0.19)	-0.89	-6.98	0.06
	Lineage 1	681	49	14	10	0.762(0.041)	0.24(0.04)	0.46(0.17)	-1.46	-2.88	0.06*
	Lineage 2	681	84	16	16	0.807(0.033)	0.37(0.02)	0.47(0.16)	-0.59	-4.32	0.08
VWF	***J. jaculus***	874	176	52	59	0.957(0.008)	0.85(0.03)	1.04(0.27)	-0.65	-29.91	0.07
	Lineage 1	874	45	31	23	0.938(0.026)	0.57(0.07)	0.83(0.28)	-1.11	-10.77	0.08*
	Lineage 2	874	131	33	37	0.933(0.012)	0.56(0.02)	0.69(0.20)	-0.70	-15.72	0.07
***VWF* without recomb**	***J. jaculus***	514	153	23	18	0.894(0.011)	0.49(0.03)	0.80(0.24)	-1.08	-4.26	0.05
	Lineage 1	516	26	6	6	0.717(0.079)	0.38(0.05)	0.30(0.15)	0.76	0.14	0.16
	Lineage 2	514	127	18	14	0.861(0.016)	0.44(0.27)	0.65(0.21)	-0.89	-2.48	0.06

L, number of sites excluding gaps; **n**, number of sequences; **S**, number of segregating sites; **H**, number of haplotypes; **Hd**, haplotype diversity; **π**, nucleotide diversity per site; $\boldsymbol{\theta_W}$, computed from the number of segregating sites; **D**, Tajima's D; $\boldsymbol{F_S}$, Fu's F_S; $\boldsymbol{R_2}$, Ramos-Onsins & Rozas's R_2. Significant values indicated *(P<0.05), **(P<0.01), ***(P<0.001).

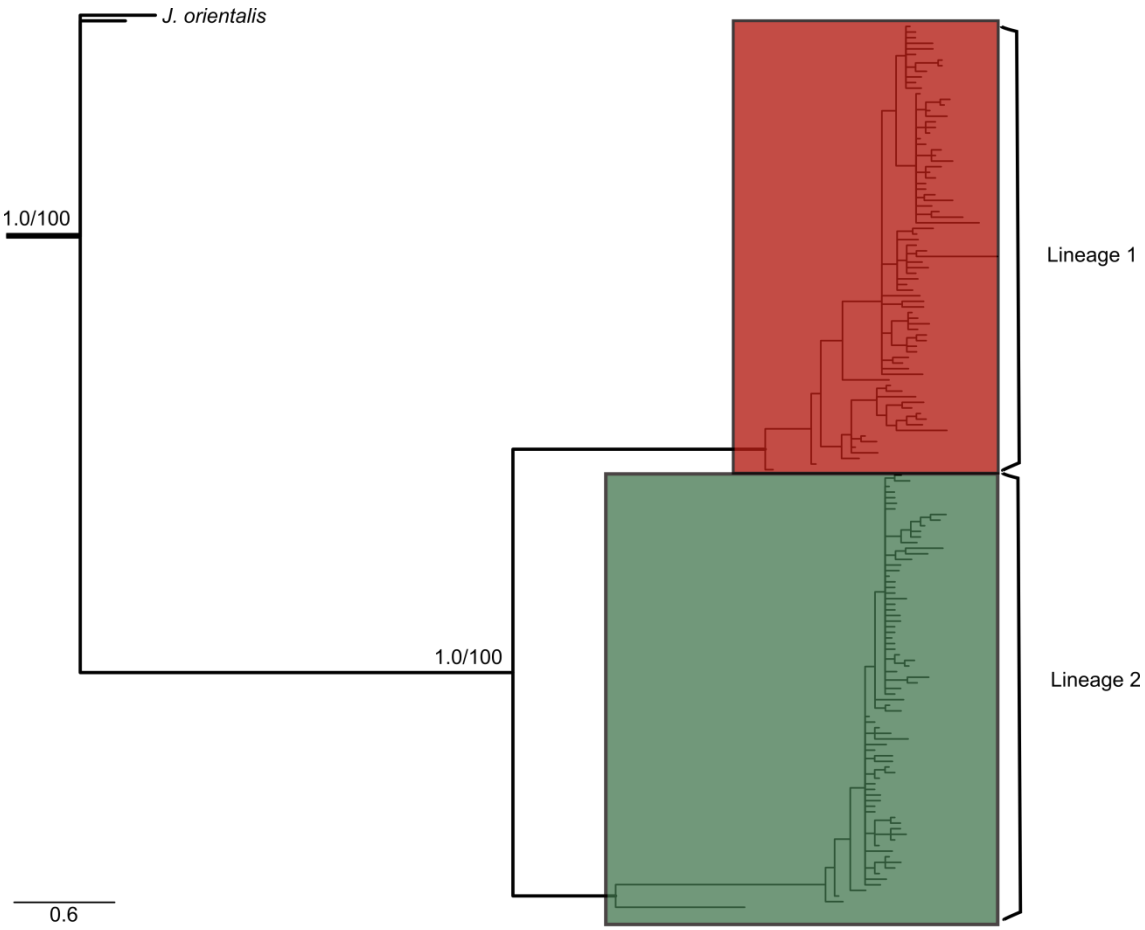


Figure 6. Phylogenetic tree based on Bayesian inference showing the relationship among the two lineages of *J. jaculus* for the long fragment of the cytochrome *b* (*cytb*) gene (n=210). Values on branches indicate posterior probability support and bootstrap values of the Bayesian and Maximum-Likelihood analysis respectively. On each clade the respective lineage is indicated (red – Lineage 1; green – Lineage 2). *J. orientalis* (n=7) was used as outgroup.

The two mitochondrial lineages can similarly be distinguished by the majority of the nuclear loci, although without consistent support values (Figures 7 and 8). The X-linked gene produced strong supports for Lineage 2, but haplotypes from Lineage 1 had a basal position in the tree (Figure 7). The phylogenetic network of the *DBX5* intron retrieved the two lineages well differentiated with 12 mutational steps separating them, wherein the size of each polymorphism is marked on the respective mutation position (Figure 8). Phylogenies of the autosomal loci recovered diversified patterns of relationship among haplotypes, generally with lower robustness when delimiting the two clades, but still both lineages can be distinguished. Contrary to that observed in *DBX5*, in the locus *ADRA2B*, haplotypes from Lineage 2 have a basal position in the tree, while Lineage 1 forms a clade, although with low support values. One haplotype

of *J. orientalis* appears clustered within *J. jaculus* haplotypes though no reliable conclusions can be made since support values for the ML approach are below 0.50 (Figure 7). Notwithstanding, on both ML and BI methods, *J. orientalis* haplotype clustered within *J. jaculus*. *υWF* analyses also recovered distinct tree topologies by BI and ML methods, which do not clarify the relation among haplotypes of Lineage 1, but Lineage 2 appears relatively well defined with high support values for both methodologies (Figure 7). Despite the diversified pattern of relationships between haplotypes, most phylogenies recover the two lineages. However, geographical structure is not evident since both lineages have the same distribution range, occurring mostly in geographic sympatry (see Figure 3 and Annex 8).

No clear evidence of gene flow was found between lineages since for the majority of the loci the two lineages were clearly separated and allele sharing was nearly absent (Figures 7 and 8). It was only detected in two exons – *GHR* and *IRBP*. For *GHR* locus one individual from Lineage 2 had a genotype where both alleles clustered within Lineage 1 (Figures 7 and 8). This individual clustered within Lineage 2 at all other loci analysed. In *IRBP* exon the opposite occurred, one individual belonging to the Lineage 1 had alleles from Lineage 2 (Figures 7 and 8). This individual grouped within Lineage 1 in all other loci analysed. These occurrences of shared variation could have been the consequence of the retention of ancestral polymorphism or recent gene flow between lineages. For the *IRBP* exon, the individual that denoted some level of allele sharing among lineages was excluded by the IMgc software, which detects and removes any traces of recombination within each locus, and so this possible instance of gene flow within this locus was not included in the final dataset. However, it was considered for *GHR* locus once this exon did not recovered any level of recombination.

Additional cases of potential recombination were identified for *DBX5* intron and *υWF* exon. From the *DBX* gene two haplotypes were discarded along with a segment of the gene (around 5% of the 5' end: 10bp). The two haplotypes detected were from Lineage 2. From the *υWF* exon a total of 19 unique haplotypes and a portion of the gene (about 41% of both ends: a total of 359bp) were eliminated. Out of the 19 unique haplotypes, 14 belonged to Lineage 1 and 5 to Lineage 2. The *IRBP* gene retrieved a set of 23 unique haplotypes where events of recombination may occur, plus 37% of the gene fragment, discarded from both ends (377bp). The 23 alleles discarded comprised 1 from Lineage 1 and 22 from Lineage 2, including the homozygous genotype that reported evidences of gene flow. The removal of these possible recombinant haplotypes in the three mentioned genes generated analogous phylogenies to the ones acquired with the whole dataset. However, some differences can be pointed: the *DBX* gene presented higher support for the node in Lineage 2, but still Lineage 1 haplotypes

did not cluster in a clade but maintained a basal position; and *ὐWF* exon recovered a distinct topology where one haplotype from Lineage 2 clustered within Lineage 1 (see Annexes 9 and 10 for the phylogenetic inferences with the non-recombinant datasets). Additionally, since *cytb* presented high levels of unique haplotypes, the data was also tested for recombination, but no significant probability was found.

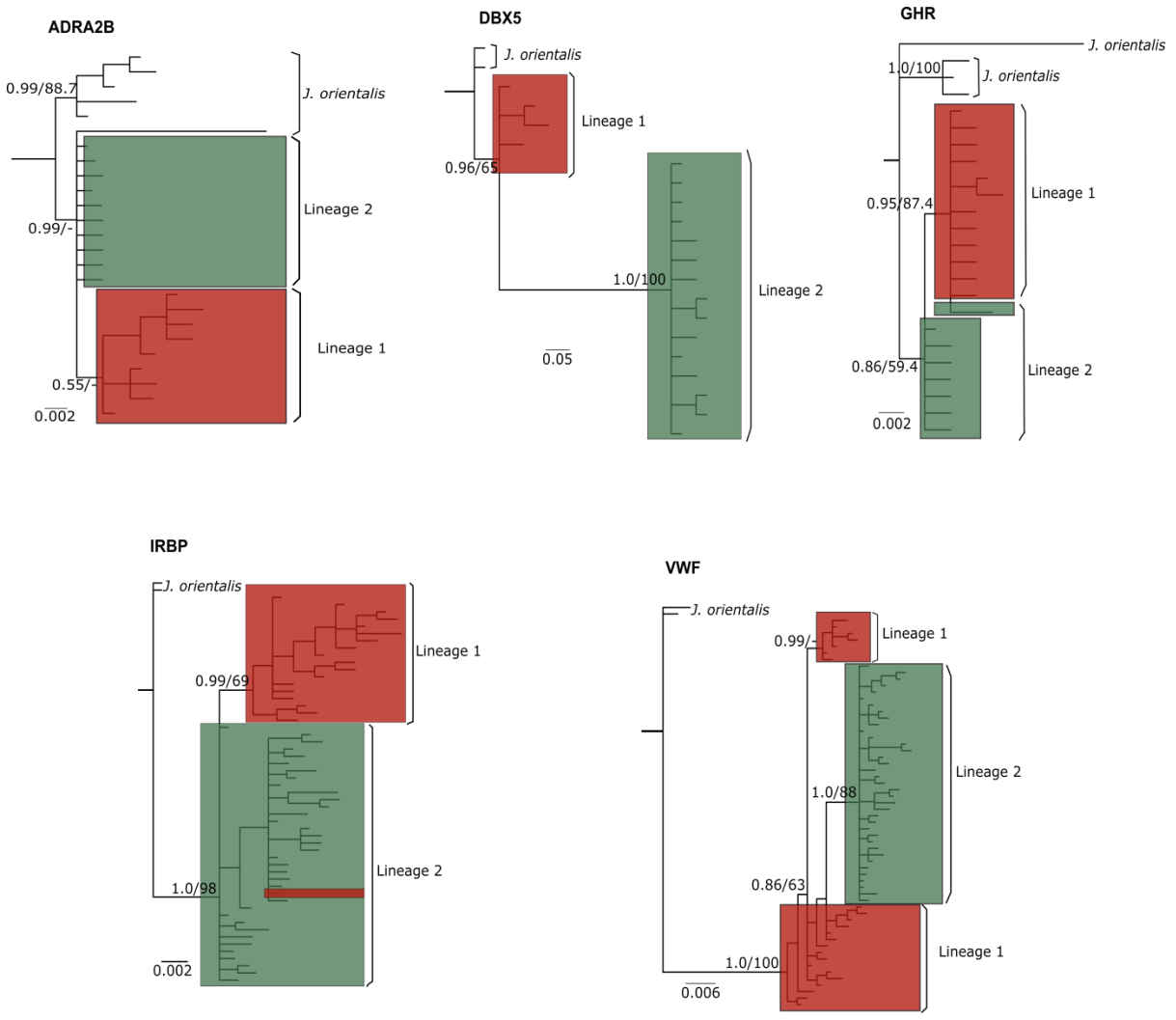


Figure 7. Phylogenetic trees based on Bayesian inference showing the relationships among *J. jaculus* specimens for the X-chromosome intron (*DBX5*) and the nuclear autosomal genes (*ADRA2B*, *IRBP*, *GHR* and *ὐWF*). Number of sequences used for each locus is specified in Table 1. Values on branches designate the posterior probability support and bootstrap values for Bayesian and Maximum-Likelihood approaches, respectively. The two colours represent the two mitochondrial lineages (red – Lineage 1; green – Lineage 2). *J. orientalis* (n=6) was used as outgroup.

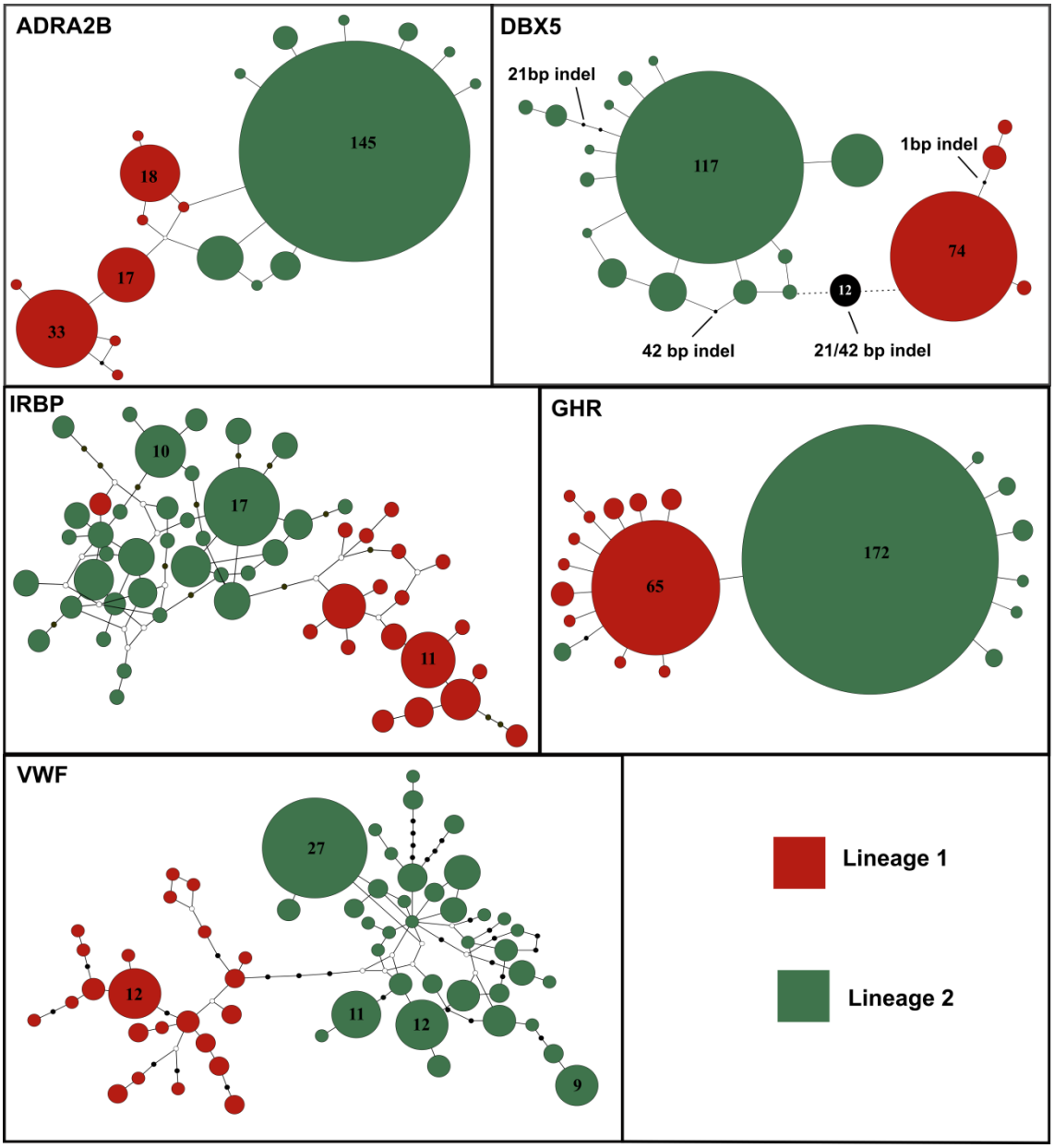


Figure 8. Statistical parsimony haplotype networks of the X-chromosome intron (*DBX5*) and nuclear autosomal genes (*ADRA2B*, *IRBP*, *GHR* and *υWF*) of *Jaculus jaculus* specimens (number of sequences used for each locus is specified in Table 1). Each circle represents one haplotype and the circle area is proportional to the frequency of each haplotype. Total frequencies are indicated for more common haplotypes. The size of the branches is proportional to the number of nucleotide differences between haplotypes, and dots on branches specify mutational steps. The insertion/deletion polymorphisms (indels) of *DBX5* were coded as single mutations (see Annex 4) and so the sizes of the indels are indicated on the respective mutational step. Due to the large number of mutational steps of *DBX5*, the number of mutational steps is indicated (12). Haplotypes are coloured according to the respective mitochondrial lineage.

Neighbour-net networks based on uncorrected patristic distances and bootstrap analysis demonstrated the two lineages as very well differentiated with both approaches (*cytb* and nuclear genes) retrieving the maximum bootstrap value (Figure 9). Nonetheless, a better resolution was obtained with *cytb* analyses (Figure 9a).

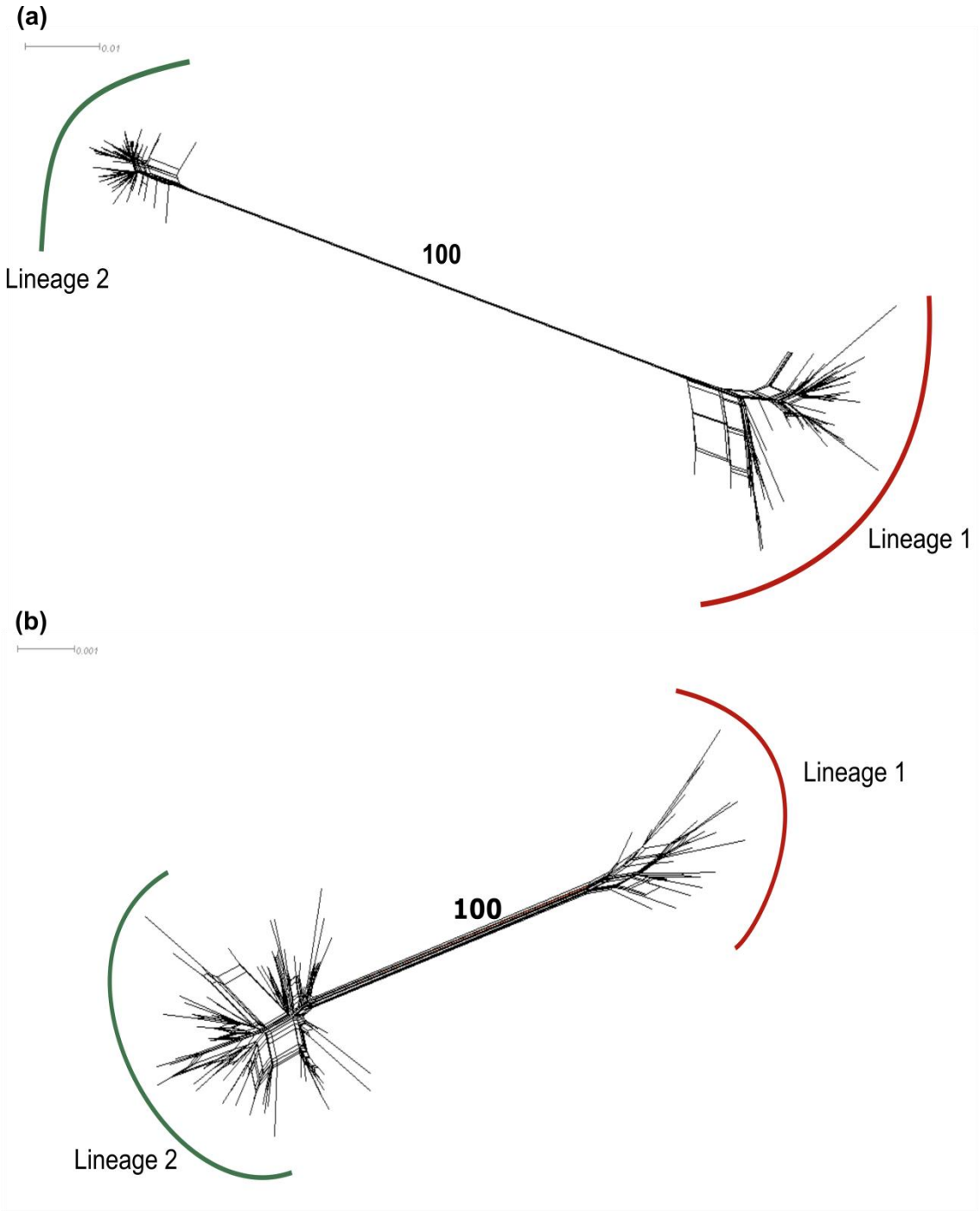


Figure 9. Neighbour-net networks based on uncorrected patristic distances as implemented in SPLITSTREE for (a) the cytochrome *b* gene (897 bp) and (b) the combined nuclear genes (*DBX5*, *ADRA2B*, *GHR*, *IRBP* and *υWF*; 3740 bp). Numbers indicate bootstrap values. Scale bar represents 1% sequence divergence. The two lineages are indicated on each clade.

3.2. Species tree inference and molecular dating

The species tree inferred by *BEAST is strongly supported, having a consistent posterior probability of 1 for the splitting nodes between lineages, and between *J. orientalis* and *J. jaculus*, for all analysis (results of the combined three runs are presented in Figure 10). By applying the universal value for *cytb* evolutionary rate of 0.02 substitutions per site/per million years (Triant & DeWoody 2006; Nabholz *et al.* 2008; Ben Faleh *et al.* 2012a), *BEAST estimations show that the splitting between *J. orientalis* and *J. jaculus* species occurred during the Late Miocene-Pliocene transition, around 5.77 million years ago [95% highest posterior density (HPD) between 4.28 and 7.32 Mya]. The split between the two lineages is estimated to have occurred through the transition of Pliocene to Pleistocene, when the major glaciations of the Quaternary have started, about 3.73 (2.67-4.81) million years before present (BP).

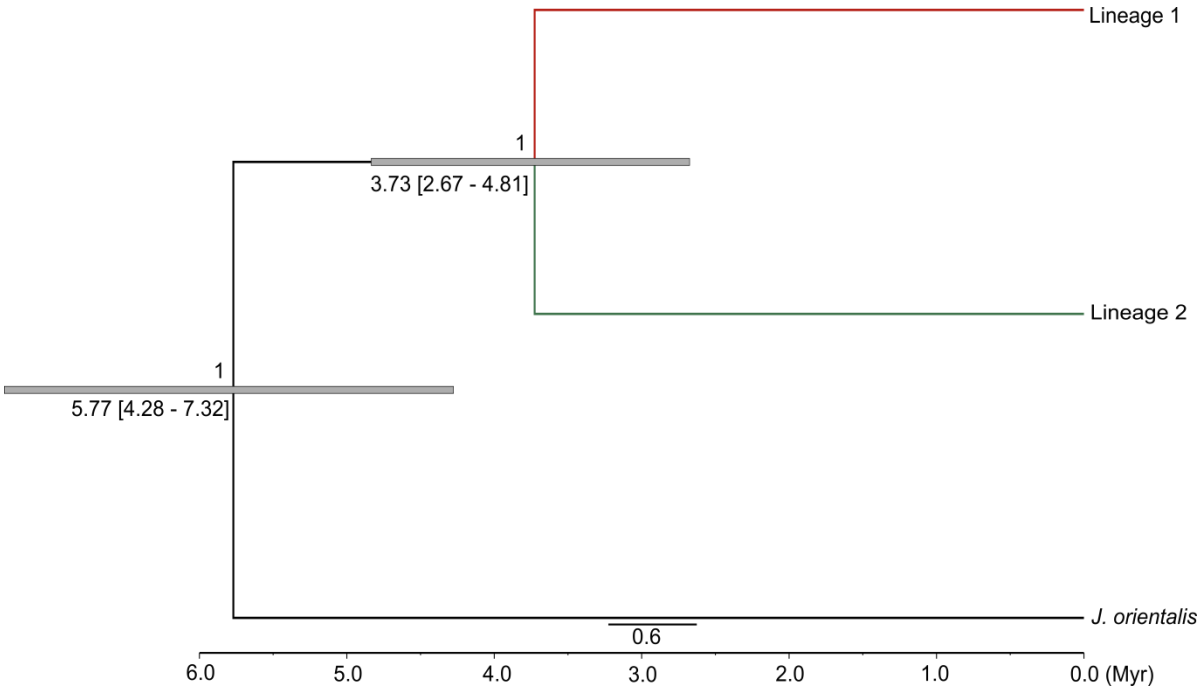


Figure 10. *BEAST species tree inference output for all loci analysed. The posterior probability of each split is shown on each node and grey bars display the 95% highest posterior density intervals for the estimated split times between the two lineages and *J. jaculus – J. orientalis,* by applying a *cytb* mutation rate of 0.02 substitutions per site per million years (values are presented below bars). The timescale in million years (Myr) is presented below the tree.

3.3. Population genetics and demographic analyses

The two lineages diverge more substantially for mtDNA than for nuclear loci, although the intron 5 from the *DBX* gene also reveals a high genetic divergence between lineages (Table 2). From the *cytb* divergence we can denote a really high genetic differentiation among lineages (10.80%), being similar to that observed between the outgroup species (*J. orientalis*) and each of the lineages (13.40% and 13.10% for Lineage 1 and 2 respectively). The DBX intron also reveals a high divergence between lineages (3.10%), being quite higher than the genetic diversity separating *J. orientalis* and Lineage 1 (0.40%) but similar with the genetic variation among Lineage 2 and outgroup specimens (3.40%). Divergence found in the autosomal loci was generally lower when compared with the variation found in mtDNA and *DBX5*, but always relatively similar with the genetic variability detected among the outgroup species and each of the lineages, with the exception of *υWF* locus that exhibited a higher differentiation between *J. orientalis* and *J. jaculus* specimens (Table 2).

Substantial variation was observed in the levels of polymorphism among lineages and among loci (Table 1). The *cytb* gene displayed the highest diversity values compared with all loci analysed, with an overall haplotype diversity (Hd) of 0.99, a nucleotide variation of 5.20% and a θ_W of 4.20%. The highest values of genetic variation for each lineage were also detected at the mitochondrial level, with a relatively lower value of nucleotide diversity for Lineage 2. In the majority of the cases, Lineage 1 showed higher values of genetic variability when compared with Lineage 2, the opposite just occurred in the locus *IRBP* (with and without recombination) and *υWF* (without recombination), wherein the polymorphism levels of Lineage 2 are considerably higher than Lineage 1 (Table 1). The *DBX* intron exhibited medium levels of haplotype diversity but relatively low values of nucleotide diversity, very similar with the ones found within *GHR* locus. Of the autosomal genes, *IRBP* and *υWF* were the ones with higher levels of polymorphism, especially haplotype diversity (Hd), being quite comparable with that found in *cytb* gene. Overall, the *GHR* exon recovered lower values of genetic diversity, particularly in nucleotide diversity. Differences between recombinant and non-recombinant datasets were more visible in *IRBP* and *υWF*, where a relevant decay of the levels of polymorphism (Hd, π, θ_W) can be observed (Table 1).

Table 2. Divergence between the two lineages described for *J. jaculus* and between each lineage and the outgroup used in phylogenetic analyses (*J. orientalis*). Additionally, the divergence (Dxy) between *J. jaculus* (combining all sequences) and *J. orientalis,* and between other closely related species of rodents were performed for comparative analysis (species and GenBank accession numbers below). The standard errors are based on 10,000 replicates (in parenthesis) and all estimates are given as percentages.

Locus		L1/L2	J. orientalis/ L1	J. orientalis /L2	*J. orientalis/ J. jaculus*	Other rodent species
Cytb	**Dxy**	10.80 (1.10)	13.40 (1.30)	13.10 (1.30)	13.10 (1.10)	13.90(1.20)[a] 8.30(0.90)[b]
	Da	9.70 (1.10)	12.3 (1.20)	12.6 (1.30)	9.90 (1.00)	-
DBX5	**Dxy**	3.10 (1.00)	0.40 (0.30)	3.40 (1.10)	2.50 (0.80)	1.10 (0.60)[c]
	Da	3.00 (1.00)	0.30 (0.30)	3.40 (1.10)	1.80 (0.60)	-
ADRA2B	**Dxy**	0.50 (0.20)	1.10 (0.30)	0.70 (0.20)	0.80 (0.20)	2.80 (0.50)[d]
	Da	0.30 (0.20)	0.60 (0.20)	0.30 (0.20)	0.40 (0.20)	-
GHR	**Dxy**	0.20 (0.10)	0.80 (0.20)	0.60 (0.20)	0.70 (0.20)	0.40 (0.20)[e]
	Da	0.10 (0.10)	0.50 (0.20)	0.30 (0.20)	0.40 (0.20)	-
IRBP	**Dxy**	0.80 (0.20)	0.80 (0.20)	0.60 (0.20)	0.30 (0.10)	0.50 (0.20)[f]
	Da	0.40 (0.10)	0.60 (0.20)	0.40 (0.20)	0.10 (0.10)	-
υWF	**Dxy**	1.30 (0.30)	3.10 (0.50)	3.40 (0.60)	3.10 (0.70)	1.40 (0.30)[g]
	Da	0.80 (0.30)	2.60 (0.50)	2.90 (0.60)	2.70 (0.70)	-

L1= Lineage 1; **L2**= Lineage 2; **Dxy**=Average number of nucleotide substitutions per site between populations (average raw DNA divergence); **Da**= Number of net nucleotide substitutions per site between populations (average net DNA divergence).

[a] *Microtus arvalis* (GQ352469) / *Microtus agrestis* (GQ352470) (Bannikova *et al.* 2009)

[b] *Microtus arvalis* (AY513809) / *Microtus kirgisorum* (AY513809) (Haynes *et al.* 2003; Jaarola *et al.* 2004; respectively)

[c] *Microtus arvalis* (JX284377) / *Microtus agrestis* (JX284376) (Paupério *et al.* 2012)

[d] *Acomys russatus* (FM162045) / *Acomys cahirinus* (FN984740) (Blanga-Kanfi *et al.* 2009; Edrey *et al.* 2012; respectively)

[e] *Allactaga bullata* (JQ347909) / *Allactaga balikunica* (KM397227) (Lebedev *et al.* 2012; Pisano *et al.* 2015; respectively)

[f] *Allactaga bullata* (JQ347929) / *Allactaga balikunica* (KM397136) (Lebedev *et al.* 2012; Pisano *et al.* 2015; respectively)

[g] *Microtus agrestis* (FM200055) / *Microtus socialis* (FM162067) (Montgelard *et al.* 2008; Blanga-Kanfi *et al.* 2009; respectively)

In general, the average number of nucleotide differences (π) was lower than the value of θ estimated from the number of segregating sites (S) (θ_W), which resulted in negative values for Tajima's *D* for the majority of the loci and for both lineages (Table 1). The same was detected for Fu's F_S statistics. Negative values of these estimators may be an indicative of selection events. In the majority of the loci analysed, significant

deviations from the neutrality tests were observed in both lineages, giving light to a possible expansion scenario. No signs of expansion were detected in *IRBP* and *υWF* since significant values were only observed for Lineage 1 in R_2 statistics. Despite the low levels of polymorphism identified in *GHR* gene, an excess of rare variants in this locus can be distinguished, which resulted in statistically significant values for all the neutrality tests. *ADRA2B* and *DBX5* presented significant values only for Lineage 2, while *cytb* displays a strong signal of expansion in both lineages for all statistics. Overall Lineage 2 shows a higher signal of expansion than Lineage 1, since it shows statistically significant values for almost all the loci. In Lineage 1 we may also infer a possible expansion but with an unclear pattern.

The Extended Bayesian Skyline plot (EBSP), representing the effective population size (N_e) through time, revealed signs of expansion in both lineages, that may have started around 200,000 years before present (BP), with an increase in population size that is going on until the present time (Figure 11a and b for Lineage 1 and 2 respectively). Though, the analysis suggests that the last demographic change may have started earlier in Lineage 1 than in Lineage 2. These findings corroborate the significant values detected in the neutrality tests for most of the loci analysed (Table 1). The estimates of contemporary population size are similar among lineages, with a moderately lower N_e detected for Lineage 2 (Figure 11).

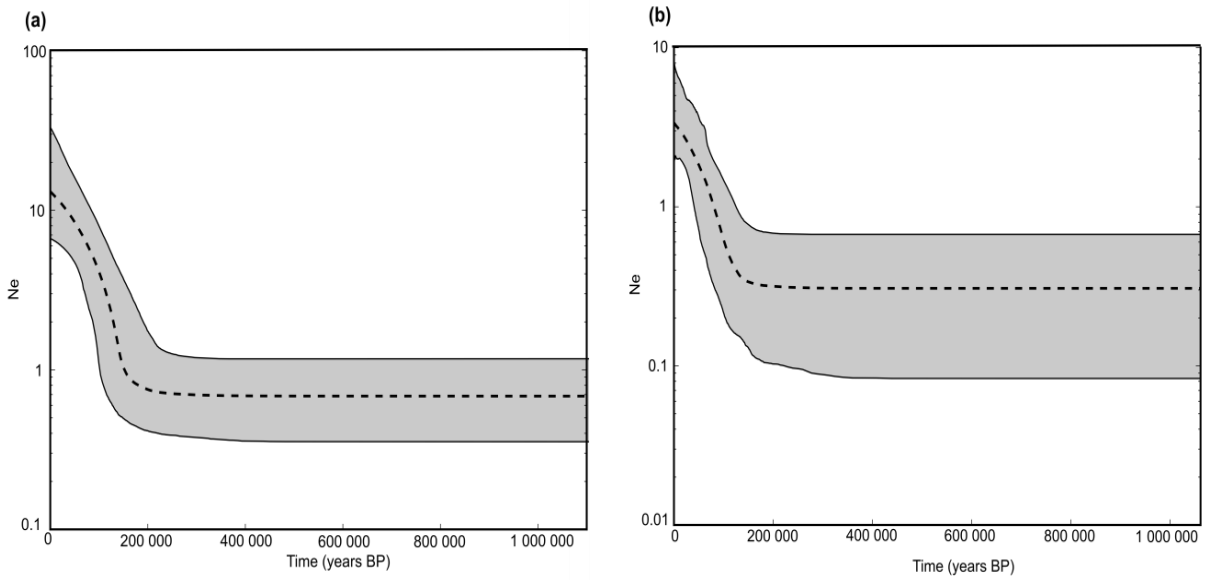


Figure 11. Extended Bayesian Skyline plots (EBSP) of the effective population size through time obtained from the three MCMC simulations. Dashed black line is the median effective population size N_e in millions, multiplied by one (mean generation time in years). Solid black lines are the 95% highest posterior density limits. The y-axis is displayed on a log scale for simplicity. (a) Lineage 1; (b) Lineage 2.

3.4. The *Isolation-with-Migration* model

Species tree inferences performed with *BEAST incorporate the uncertainty associated with the coalescent process while estimating the phylogeny, however, it does not assume the possibility of occurrence of gene flow after the initial split. In order to assess the potential impact of gene flow on the evolution of the mitochondrial and nuclear genes, our data was analysed under the IM model (Hey & Nielsen 2004). The resulting evaluations of the posterior density curves of the model parameters were consistent across independent runs. Nonetheless, the right tail of the posterior density curves was not achieved for some parameters, particularly in the estimates of divergence time and ancestral effective population size, since it failed to reach zero in both runs. The average nucleotide divergence (Dxy) and the net nucleotide divergence (Da) between *J. jaculus* and *J. orientalis* were applied as two different approaches to estimate the specific mutation rate for each locus, combined with the divergence in million years (Myr) proposed by Pisano *et al.* (2015). The two approaches were applied so that a more robust range of estimations could be achieved. Estimations of the geometric mean of the mutation rate of the 6 loci used were different in both methods applied (summary of the calculations in Annex 11) resulting in different estimates of the effective population sizes (N_e) and divergence times between lineages (t) (Table 3). Evaluations of the effective population sizes detected a higher N_e for Lineage 1. Regarding divergence times, with the Dxy approach the split between Lineage 1 and 2 is estimated to be about 3.42 (1.66-8.76) Myr before present (BP), whereas Da gives an approximation around 5.72 (2.78-14.68) Myr BP (Table 3). Population migration rates were found to be significant, wherein a higher rate was detected for $2Nm_2$, that is the rate at which genes of the Lineage 2 are replaced by genes from Lineage 1 (Table 3 and Figure 12). IMfig shows an approximate divergence time between the two lineages and their ancestral population of 6.80 (4.42-21.38) Myr BP (Figure 12).

Table 3. Maximum-Likelihood estimates (95% posterior density intervals in parentheses) of demographic parameters obtained with IMa2 between the two mitochondrial lineages (mean generation time of 1 year). The values are in millions.

	N_{e1}	N_{e2}	N_A	t	$2Nm_1$	$2Nm_2$
Dxy approach	1.22 (0.92-1.60)	0.86 (0.66-1.11)	0.75 (0.51-1.83)	3.42 (1.66-8.76)	0.14*** (0–0.38)	0.24*** (0.04-0.43)
Da approach	2.04 (1.55-2.68)	1.44 (1.10-1.86)	1.26 (0.86-3.06)	5.72 (2.78-14.68)	0.14*** (0-0.38)	0.24*** (0.04-0.43)

N_{e1}, effective population size of Lineage 1; N_{e2}, effective population size of Lineage 2; N_A, effective population size of the ancestral population; **t**, time in generations since the split of Lineage 1 and 2; $2Nm_1$, population migration rate into Lineage 1; $2Nm_2$, population migration rate into Lineage 2. Significant values indicated ***(P<0.001).

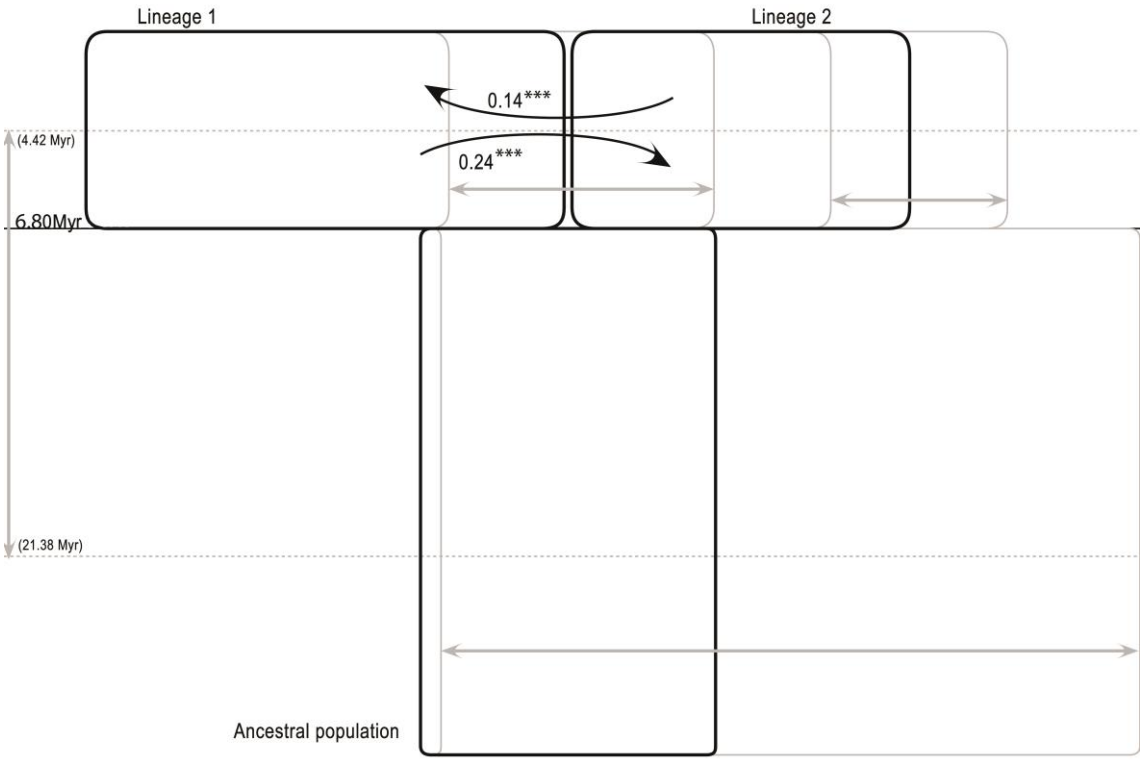


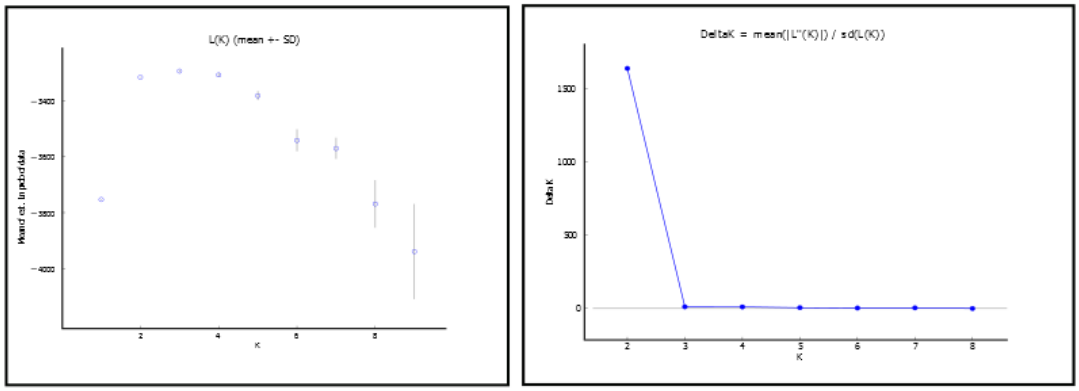
Figure 12. Phylogenetic tree obtained with IMfig wherein the boxes denote the respective population, horizontal lines represent the splitting times and the curved arrows indicate migration. Time is represented as depth in vertical axis, where the split time among the two lineages and the ancestral population is represented (6.80 Myr). The most recent time is on top where the sampled populations are (Lineage 1 and Lineage 2). The 95% Highest Posterior Density (HPD) intervals for splitting time is given by dashed lines and double-headed arrows in the vertical axis (4.42-21.38 Myr). Population size is represented as width through the horizontal axis. The 95% HPD for a population size is given both by a double headed arrow and by faint boxes. Significant values are indicated ***(P<0.001).

3.2. Microsatellite analyses

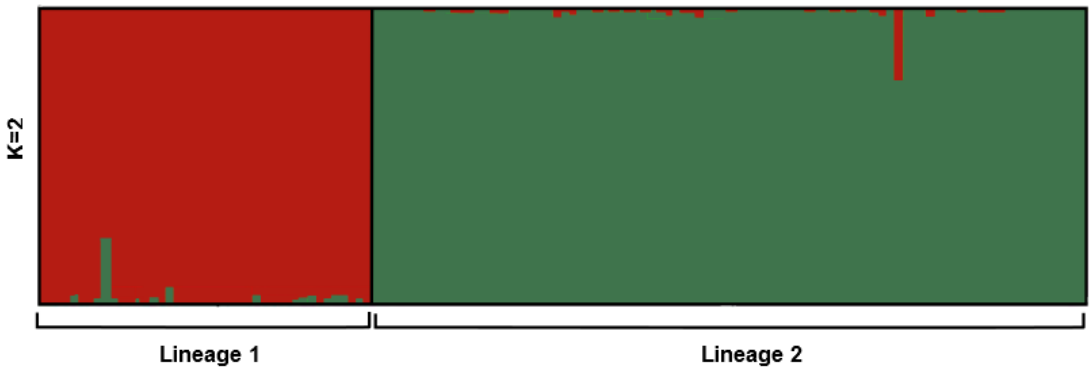
Microsatellite analyses were conducted considering the two lineages as populations in order to determine the current population structure, as well as to clarify possible levels of gene flow among the putative species. Assessments of population structure were performed for the complete set of the field samples (152) wherein 148 samples were effectively amplified. A total of 132 specimens was analysed for 6 loci after a removal of 40% of missing data and all markers in which significant deviations to HWE and LD assumptions were observed (Jac04, Jac07, Jac11, Jac12, Jac24, and Jac27).

Structure Harvester results highlighted K=2 as the most likely number of clusters which best explain our dataset, both for DeltaK and L(K) methods (Figure 13a). Structure bar plot (Figure 13b) exhibited a clear separation between Lineage 1 and Lineage 2, as revealed for mitochondrial and nuclear loci analyses. The average proportion of membership for each cluster was Q_1=98.5% for Lineage 1 (in red in Figure 13b) and Q_2=98.9% for Lineage 2 (in green), supporting the separation between groups. Individual proportions of memberships can be consulted in Annex 12. Two individuals can be considered as outliers since their proportion of membership deviate from the one expected. One individual from Lineage 1 exhibited 79.50% of assignment to its respective cluster and 20.50% to Lineage 2; the second individual belongs to Lineage 2 and reveals a proportion of membership to its own cluster of 75.40% and 24.60% to Lineage 1 (see Annex 12). These two specimens can be distinguished in the Structure plots (Figure 13b). The Principal Coordinate Analysis (PCA; Figure 13c) based on the genetic distances suggested a moderate genetic distance between clusters, with Axis 1 revealing 16.53% of genetic variability between the two lineages. Axis 2 denotes 5.30% of genetic variation, where a higher variability within Lineage 2 is perceived, although with no geographical basis. Two individuals reveal high similarity with the opposite lineage, which is not corroborated by the Structure analysis or phylogenetic inferences. Furthermore, the two specimens for which the genotypes clustered with the opposite lineage in *GHR* and *IRBP* markers (see Figures 7 and 8) were not used in microsatellite analyses due to its high percentage of missing data, and so further conclusions could not be inferred. Since a high variability within each lineage was detected, additional structure bar plots were analysed for K=3 and K=4, where a higher variability is observed for Lineage 2 (Annex 13).

(a) Structure Harvester



(b) Structure



(c) PCA

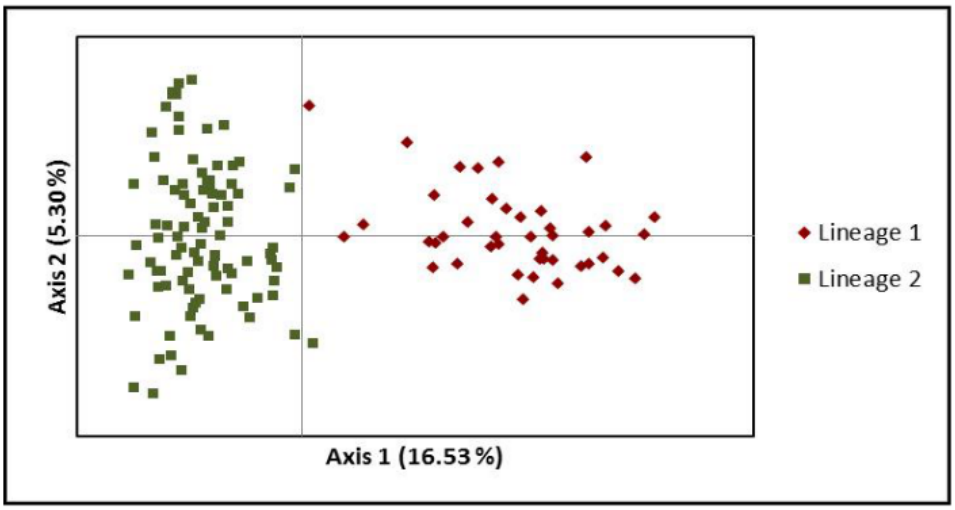


Figure 13. Population structure analyses of 132 *J. jaculus* specimens (42 from Lineage 1 and 90 from Lineage 2) based on 6 microsatellite loci. (a) Structure Harvester graphic output of Delta K and Mean L(K); (b) Structure bar plot of Bayesian assignments of individual to the respective cluster (K=2). Vertical bars indicate individuals and the colours within each bar correspond to the probability of membership of each specimen to a cluster (In red – Lineage 1; in green – Lineage 2). (c) Principal coordinate analysis (PCA) based on the individual-by-individual genetic distances.

High levels of polymorphism were detected between and within each cluster (see genetic indexes summarized in Table 4). The same number of alleles (Na) and identical numbers of effective alleles (Ne) were found in both lineages. The microsatellite markers can be considered as highly polymorphic due to the notorious number of alleles observed in only six markers (12.83), varying from 9, in marker Jac16, to 29 alleles, in Jac37 (see number of alleles sampled per locus and population in Annex 14). Relatively high levels of heterozygosity (Ho) were detected in each lineage and the whole population sampled (74.70%), wherein Lineage 1 appears to have higher values for genetic diversity (80%) than Lineage 2 (69.40%). No significant deviations were found among values of observed (Ho), expected (He) and unbiased expected heterozygosity (uHe).

Differentiation measures across the two populations recovered moderate values for the Total Fixation Index (F_{IT}) and the Fixation Index (F_{ST}), though the two values appear to be statistically significant (P<0.05) (Table 4). These values imply a relatively high level of differentiation between lineages.

Table 4. Mean Heterozygosity, F-statistics and Polymorphism data for both lineages of *J. jaculus* based on the analysis of microsatellite loci.

Cluster		N	Na	Ne	Ho	He	uHe	F_{IS}	F_{IT}	F_{ST}
Lineage 1	**Mean**	42,000	12.833	5.740	0.800	0.808	0.818	-	-	-
	SE	0,000	1.376	0.809	0.035	0.026	0.027			
Lineage 2	**Mean**	90,000	12.833	5.865	0.694	0.702	0.706	-	-	-
	SE	0,000	2.441	1.392	0.133	0.135	0.136			
Total	**Mean**	66,000	12.833	5.803	0.747	0.755	0.762	0.022	0.176(*)	0.161(*)
	SE	0,000	1.336	0.768	0.068	0.068	0.068	0.011	0.029	0.032

SE – Standard error; **N** - Sample Size; **Na** – Number of Alleles; **Ne** - Number of Effective Alleles; **Ho** - Observed Heterozygosity; **He** - Expected Heterozygosity; **Uhe** - Unbiased Expected Heterozygosity; F_{IS}- Inbreeding Coefficient; F_{IT} – Total Fixation Index ; F_{ST} – Fixation Index. Significant values indicated *(P<0.5).

4. Discussion

The comprehensive approach of the present study complements previous assessments of the phylogenetic relationships between the two *Jaculus* lineages, allowing an almost full review of the phylogenetic pattern and evolutionary history of the species. By analysing different types of molecular markers, we shed light to the evolutionary processes behind the differentiation of the putative species, as well as on the underlying mode of speciation. Our study provides new insights on the population history and structure of the two mitochondrial lineages, offering the first appraisal of the levels of gene flow between them, thus enabling a better inference of a reproductive isolation hypothesis.

4.1. Two phylogenetic lineages within the Lesser Egyptian Jerboa

This study clarified the phylogenetic relationship between the two lineages within *J. jaculus* species, presenting widespread and overlapped distributions across North Africa (see Figure 3). The phylogenetic inferences of mitochondrial DNA revealed two well defined and strongly supported clades (see Figure 6 and Annex 7 and 8), thus confirming the two mitochondrial lineages identified by previous studies (Ben Faleh *et al.* 2012; Boratyński *et al.* 2012, 2014a). The two lineages, that were further recovered with an autosomal marker (*vWF*; Boratynski *et al.* 2012), were also distinguished for the X-linked locus (*DBX5*) and three additional autosomal genes (*ADRA2B*, *GHR* and *IRBP*; see Figures 7 and 8), although without consistent support values. Therefore, the six loci analysed appear to reflect a genome-wide outcome since they are generally congruent in the support of the two lineages.

However, there are some differences in the topology recovered for the individual gene trees and the support values for the lineages within them. Overall, the autosomal markers, compared to *cytb*, retrieved lower support values when delimiting the two clades, showing diversified patterns of relationships among specimens. It is known that the configurations found in individual gene trees are sometimes distinct from what is detected in the species tree, due to the retention and incomplete lineage sorting of ancestral polymorphisms and/or introgression, particularly in the case of recently diverged species (Avise *et al.* 1983; Tajima 1983; Pamilo & Nei 1988; Maddison 1997). Moreover, the heterogeneity of the branch lengths is also believed to counteract to the

consistency of the phylogenetic signal in the separate gene trees (Edwards 2009). This is especially common in closely related species with coincident distributions ranges (Carling & Brumfield 2008). Notwithstanding, when we applied the newly developed coalescent method that considers incomplete lineage sorting as the main driver of incongruity among genes (Heled & Drummond 2010), it produced concordant results with the ones found in mitochondrial analyses and in the majority of the nuclear loci, recovering the two lineages well differentiated (see Figure 10).

Although there are evidences of the existence of two distinct evolutionary units, the fact that their distributions ranges overlap throughout all their geographic extent, increases the probability of gene flow between them. Two individuals were identified with different lineage assignments in two autosomal genes (*GHR* and *IRBP*; Figures 7 and 8), wherein a potential recombinant genotype was recognized in *IRBP*. The two specimens occur in the regions of Western Sahara, along the Atlantic coast, and in Mauritania, where both populations appear to be in the same geographical area (see Figure 3). These results were followed by IM analyses, where both lineages appear differentiated with significant levels of gene flow between them (see Table 3; Figure 12). The same pattern was recovered in the contemporary genetic structure inferred by microsatellite data, retrieving two differentiated clusters with negligible levels of allele admixture (see Figure 13; Table 4). However, no clear contact zones can be defined since both lineages coexist in geographical sympatry throughout Northern Africa extent.

4.2. Insights into the evolutionary history of *Jaculus* species

The radiation of jerboas has been placed in Asia since the Late Miocene, being related with the development of open habitats with an increased aridity (Shenbrot *et al.* 2012; Zhang *et al.* 2012; Pisano *et al.* 2015). Phylogeographical studies on a wide range of desert organisms [from animals (Nicolas *et al.* 2009) to plants (Besnard *et al.* 2007)] often link the origin of genetic variation to the expansion phases of the arid environments, as a consequence of the climate changes that occurred between the periods of Miocene and Pleistocene (Kimura 1980; Ruddiman *et al.* 1989; Griffin 2002; Rato *et al.* 2007; Wagner *et al.* 2011; Gonçalves *et al.* 2012). The Late Miocene, particularly between 8 and 5 Myr BP, was a crucial period for the uplift of the Tibetan Plateau, which along with the concurrent development of the monsoon led to an increased aridity in Central Asia (Wang *et al.* 1999; An *et al.* 2001; Zheng 2004). These assumptions were supported by our species tree inference, estimating a divergence

time between *J. jaculus* and *J. orientalis* during the transition of Late Miocene to Pliocene, around 5.77 (4.28 – 7.32) Myr BP (Figure 10). The most recent common ancestor (MRCA) of *J. jaculus* might then have colonized a wide region from Central Asia to North Africa, giving rise to *J. jaculus* species expansion across the Northern Africa extent through vicariance events (around 5.15 Myr BP; Pisano *et al.* 2015).

During the Mid-Late Pliocene, the North African regions went through intense fluctuations in space and time, shifting between more humid and more arid phases (Rognon 1993; Foley *et al.* 2003; Kröpelin *et al.* 2008; Drake *et al.* 2011). These shifts in climate and land-cover triggered significant alterations in the Sahara-Sahel boundaries, leading to periodic modifications of desert biota (Douady *et al.* 2003; Brito *et al.* 2014). Our species tree suggest a divergence time between the two lineages along the Late Pliocene/Early Pleistocene boundary, about 3.73 (2.67 – 4.81) Myr BP (Figure 10), what is roughly consistent with the estimates obtained with IM, using the nucleotide divergence (Dxy) estimator [3.42 (1.65-8.76) Myr; Table 3]. The divergence time acquired with estimations of the net nucleotide divergence (Da) denotes for an ancient split of the two lineages [5.72 (2.77-14.68) Myr; Table 3], which was expected since this parameter does not consider the within species divergence (Wakeley 1994), and so it may actually refer to the time of divergence between *J. orientalis* and *J. jaculus*, which is similar to the splitting time obtained in the species tree inference (Figure 10). Hence, the two putative species may have suffered subdivisions as a consequence of the fluctuations across the Northern African range, maintaining relative low effective population sizes (Figure 11), probably influenced by the recurrent humid phases, which are believed to counteract expansion events in xeric species (Brito *et al.* 2014). The within species divergence might have occurred during the following periods of intense aridification triggered by the Quaternary climatic alterations. Our estimates indicate an older divergence of *Jaculus* species when compared to other rodent species such as *Acomys* [1.25 (0.65–1.94) Myr; Nicolas *et al.* 2009] or *Mastomys* [2.82 (1.61–4.20) Myr; Brouat *et al.* 2009]. Nevertheless, the dating estimates are very imprecise due to the lack of attempts in appraising the substitution rate in these rodents, and the unavailability of the required fossil records to produce robust estimations.

Previous assessments on the historical demography of the species indicated potential signs of expansion in the two lineages, with two substantially different evolutionary histories (Boratyński *et al.* 2012). Our results corroborate these suggestions, with some detected differences between lineages, regarding times of

expansion and colonization events. The recovered dissimilarities may be a consequence of the different molecular methods used in this study, highlighting the importance of a comprehensive approach in assessing the evolutionary history of a species. For the majority of the loci analysed both lineages recovered significant values for the neutrality tests, thus rejecting a model of population at equilibrium (Table 1). From these analyses, a clearer pattern of expansion is observed for Lineage 2, since it showed a higher significance across loci for Tajima's *D* and Fu's F_S estimators than Lineage 1 (Table 1). EBSP analyses gave concordant results in exhibiting a clear sign of expansion in both populations, though a more evident expansion signal is recovered for Lineage 1 with more than a ten-fold increase in population size (Figure 11a). The two putative species apparently started experiencing a substantial expansion around 200,000 years BP, suggesting a slightly earlier expansion for Lineage 1 (Figure 11). This period coincides with the major climatic oscillations of the Upper Pleistocene when North African regions experienced alternate humid and arid phases, inducing critical changes on the genetic signature of several species (Hewitt 1996; Guiller *et al.* 2001; Cosson *et al.* 2005; Guiller & Madec 2010). There are evidences that the Saharan climate fluctuations were correlated with the dramatic glaciations in the northern hemisphere (Raymo 1994; Le Houérou 1997; deMenocal 2004), during which large portions of the African range were characterized by dryer climates, thus restricting savannas and woodlands in favour of more desert-like habitats (Van Zinderen 1978; Dupont *et al.* 2000; deMenocal 2004; Cowling *et al.* 2008). Climate changes of such magnitude are known to play crucial roles in modelling species distributions and could have stimulated the population expansions of *Jaculus* species. These results are in agreement with those found for other West African rodents, such as *Acomys* (Nicolas *et al.* 2009), and are consistent with a meta-analysis clarifying the role of Pleistocene conditions in shaping the phylogeographical patterns of vertebrate species (Avise *et al.* 1998).

4.3. Assessing the processes behind speciation

Our results were congruent in defining two highly divergent lineages which may have undergone a substantial expansion since the Late Pleistocene period. Demographic analyses detected a higher effective population size for Lineage 1 (Figures 11 and 12; Table 3), suggesting an enhanced performance in occupying wider ranges when compared to Lineage 2 (Boratyński *et al.* 2012, 2014a). However, both

lineages are found often in sympatry throughout the North African extent with no apparent geographic structure between them, making it difficult to identify the processes behind their differentiation.

The reconstructed networks showed a lack of clear geographical structuring of the two putative species (Figures 3 and 8; Annex 8), suggesting relatively long distance migrations between remote locations. Similarly, the genetic variability detected within each population with microsatellite data, particularly in Lineage 2 (Figure 13c; Annex 13), is not supporting a clear geographic pattern thus implying a lack of putative barriers to gene flow within North Africa. Nevertheless, our limited geographic cover of the species distribution and the small sampling size by region/locality may under-represent the genetic diversity of *J. jaculus* species.

Previous studies suggested that the putative species may have evolved reproductive isolation due to the lack of gene exchanging in mitochondrial and nuclear markers (Ben Faleh *et al.* 2012a; Boratyński *et al.* 2012, 2014a). Our results gave a new perspective to this hypothesis by revealing rough levels of gene flow between lineages. There were some evidences of gene flow or incomplete lineage sorting in the analyses of nuclear gene trees since two autosomal loci recovered some inconsistencies relative to the species tree (Figures 7 and 10). By analysing the data under a model that assumes genetic drift as the key role causing population differentiation, it is more likely to accept an introgression hypothesis (Table 3; Figure 12). IM analysis among the two putative species suggested gene flow in both directions, denoting higher levels towards Lineage 2 (Table 3; Figure 12). Despite significant, the inferred values of migration parameters were relatively low when compared with those inferred between other mammals subspecies (e.g., Won & Hey 2005; Geraldes *et al.* 2008; Bonhomme *et al.* 2009; Carneiro *et al.* 2009; Hey 2009; Stevison & Kohn 2009), but higher than the ones estimated among hares (Melo-Ferreira *et al.* 2012). It is nonetheless important to highlight that these results might be overestimated due the existence of a third mitochondrial clade, confined to the Middle East, potentially closely-related with Lineage 2, which was not sampled in the present study (Ben Faleh *et al.* 2012b). By analysing the data under the IM model with a potential third species exchanging genes one of the assumptions of this method is violated (Nielsen & Wakeley 2001; Hey & Nielsen 2004). However, for reasonable to low levels of gene flow ($N_e m$ < 0.2, similar to our estimates – Table 3), a minor impact of an unsampled third species on the estimates of the demographic parameters is recognized (Strasburg & Rieseberg 2010). In a more recent timeline, microsatellite data

shed light to the assumption of recent gene flow, by indicating potential allele exchange among populations, although without significant admixture since the majority of the individuals revealed a high membership probability to the respective mitochondrial lineage (Figure 13; Annex 12).

Notwithstanding, it is likely that the results obtained with IM analyses allow the rejection of a model of allopatric speciation with no gene flow. Therefore, we may imply rough levels of gene exchange since the divergence time of the two lineages, notably in the direction of Lineage 2. Population divergence in the presence of gene flow is often referred as evidence that local adaptation is the crucial driver of differentiation between two or more populations (Millicent & Thoday 1960; Smith 1966; Endler 1977). Indeed, previous studies have suggested that, despite the coexistence of the two lineages in geographical sympatry across the entire Sahara, they might segregate into distinct micro-habitats (Gharaibeh 1997; Boratyński *et al.* 2012, 2014a). A persistent habitat-phenotype covariation was formerly documented, proposing that natural selection may be triggering cryptic coloration of the dorsal fur, which may have induced the phenotypic divergence between the putative species. Thus, the sympatric clades might persist in ecological separation within the admixture of sandy (lighter) and rocky (darker) micro-habitats over North Africa, wherein Lineage 1 is associated with brighter and sandy areas, and Lineage 2 to the darker rocky substrates (Boratyński *et al.* 2014a). Their results revealed a tighter micro-habitat preference for Lineage 1, suggesting that Lineage 2 may be competitively excluded from its optimum conditions, which might explain the lower effective population sizes observed.

Bearing in mind these outcomes, under a spatial context, we hypothesize that a process of parapatric speciation might be involved in the diversification of the two lineages, wherein natural selection may be driving the cryptic divergence between them. However, detailed analyses regarding levels of gene flow between species are required so that we can fully distinguish whether sympatric or allopatric speciation or local adaptation had the key role in driving the observed differentiation. The low levels of gene flow found between lineages suggest that the putative species may be under a strong, although incomplete, reproductive isolation scenario.

4.4. From lineages to species

Several studies have been proposing the emergence of two putative species within the Lesser Egyptian Jerboa in the North African range, based on the

morphological, ecological and genetic diversity observed between the two lineages (Peter 1961; Ranck 1968; Gharaibeh 1997; Ben Faleh *et al.* 2010a, b, 2012b; Boratyński *et al.* 2012, 2014a). The comprehensive approach applied in the present study confirmed this hypothesis by revealing the co-existence of two putative species showing high levels of genetic differentiation, both at nuclear and mitochondrial levels.

The average *cytb* nucleotide divergence found among lineages (10.80%) was slightly higher than what was previously documented for these species (10.50%, Ben Faleh *et al.* 2012b; 10.60%, Boratyński *et al.* 2012), and notably beyond of that recovered from intraspecific studies of rodents (average 2.09% within a range of 0.00-6.29; Bradley & Baker 2001; Ben Faleh *et al.* 2010a; Paupério *et al.* 2012). Moreover, the observed divergence is above the average genetic distance generally observed between sister rodent species (mean of 9.55%, range 2.70-19.23; Bradley & Baker 2001; Baker & Bradley 2006). Compared to *Microtus* species, the values of divergence between the two lineages were significantly higher than that observed among *M. arvalis* and *M. kirgisorum*, two identified closely-related species, but lower to that detected between *M. arvalis* and *M. agrestis* (Table 2), which refer to quite distantly-related species (Jaarola *et al.* 2004; Edrey *et al.* 2012).

The nucleotide distance found within nuclear genes was much lower (Table 2), as expected for slowly evolving markers (Zhang & Hewitt 2003). In the autosomal genes, the genetic divergence detected between the two lineages was mostly inferior to that acquired for other sister species of rodents (Table 2), though the values were fairly similar to the differences between each lineage and *J. orientalis*. Furthermore, the genetic divergence revealed in *IRBP* gene was considerably higher among the two lineages than what is observed between *Allactaga balikunica* and *A. bullata*, two relatively distantly-related species (Pisano *et al.* 2015). In contrast, the X-linked gene is significantly more diverse, displaying a substantial differentiation between lineages when compared to other rodent species (Table 2). The observed genetic diversity may be a consequence of faster lineage sorting on the X-chromosome, which was expected due the smaller effective population size, representing a three-quarter of the autosomes; and could have been driven by positive selection, being associated with genetic hitchhiking (Begun & Whitley 2000; Schaffner 2004). The great differentiation might also reflect the low levels of gene flow occurring on the *DBX* gene, which is consistent with our results (Figures 7 and 8).

The observed values of genetic differentiation within *J. jaculus*, propose that the differences between Lineage 1 and Lineage 2 might relate to the diversification of

closely-related species. In fact, the observed values of genetic divergence combined with the previously proposed morphological and ecological differentiation between the two lineages, could easily meet the expectations of the "unified species concept" proposed by De Queiroz (2005, 2007), and with the integrated framework developed by Naomi (2011). Nonetheless, as discussed above, some levels of gene exchange among lineages were inferred, and so the process of reproductive isolation might be incomplete. Despite the observed differentiation related to the dorsal fur colour, the phenotypic variation found between lineages is rather continuous than dichotomous. Therefore, the admixture of habitat and geographic ranges is followed by a highly overlapping phenotypic variation between species, thus implying the evolution of cryptic diversity (Boratyński *et al.* 2014a). Under these circumstances, we may postulate the emergence of two closely-related rodent species through a process of cryptic speciation. Thereby, following that previously suggested in the literature (Peter 1961; Ranck 1968; Gharaibeh 1997; Ben Faleh *et al.* 2010a, b, 2012b; Boratyński *et al.* 2012, 2014a), Lineage 1 can be assigned to *Jaculus jaculus* (Linnaeus 1758), and Lineage 2 as *Jaculus deserti* (Loche 1867). However, due to the controversies identified in the literature (Harrison 1978; Corbet 1978; Wilson & Reeder 2005), a taxonomic key to define the species is not available. Future studies addressing the possible geographic and ecological mechanisms behind the evolution and speciation in *Jaculus* species would give basis for a clarification of the species status, which would have substantial implications for the field identification of these two species. Moreover, the limited information available regarding the species biology and geographic distribution on "The IUCN Red List" (IUCN 2008) would require a review since it could have substantial implications on the conservation status of the species.

5. Concluding remarks and future research

The comprehensive approach applied in this study validated the previous hypothesis of the co-existence of two closely-related species recognized within the Lesser Egyptian Jerboa. Novel insights regarding population history and structure of *J. jaculus* and *J. deserti* were provided, giving light in the sense of a reproductive isolation hypothesis. Results reveal low levels of gene flow among lineages, thus suggesting a strong signal of reproductive isolation, but probably incomplete, between species. Phylogenetic results imply a divergence level comparable to the one obtained between other rodent species, with the splitting age coinciding with the major climate shifts in North African regions during the Late Pliocene/Early Pleistocene boundary. Moreover, the genetic variation found within lineages suggested divergent demographic histories in the absence of clear geographic structuring, in which one of the lineages experienced expansion relatively recently and showed a smaller effective population size when compared to the other. Therefore, it is proposed that the process of speciation occurred in the presence of gene flow, wherein local adaptation probably had the key role in enhancing the recovered genetic diversity between species. However, additional analyses are required to further evaluate the possible geographic and ecological mechanisms behind evolution and speciation in the jerboas.

This study provides the first set of microsatellites markers developed for *J. jaculus*, which revealed to be a valuable tool to estimate the current population structure of the species. However, a great proportion of the developed markers exhibited significant signs of linkage disequilibrium, what constraint the robustness of the analysis. Therefore, additional microsatellite development would be advisable for further population genetic analyses. Moreover, a higher sampling effort and a more pronounced sample size per region/locality is required in order to extend the current genetic structure analyses to the whole distribution range of the species. The pronounced proportion detected of microsatellite markers with significant linkage disequilibrium would require further studies (e.g. using genomic methods) to assess the mechanisms underlying the observed connectivity among markers.

The acquired genetic data must be followed by ecological and ecophysiological studies so that a profounder knowledge of the habitat selection patterns in relation with local adaptation can be achieved. Additionally, ecological modelling would be an important tool to identify the correlation of the habitat and environmental variables with their geographic distribution.

Genome scans by means of sequencing restriction-site-associated DNA (RADs, e.g. Hohenlohe *et al.* 2010), genotype-by-sequencing (GBS, e.g. Narum *et al.* 2013), SNP genotyping (e.g. Neafsey *et al.* 2010) or even full-genome sequencing (e.g. Jones *et al.* 2012) are thought to be a valuable tool in the identification of loci with potential signs of adaptation and/or reproductive isolation. Thus, the application of these methods would enable a reliable estimation of the evolutionary forces involved in the speciation process of *Jaculus* species.

The present study contributed to improve the knowledge about the processes behind the differentiation of the two species; however, further studies are necessary to effectively validate the level of reproductive isolation between species, as well as to assess the local forces driving the intraspecific genetic diversity observed.

6. References

Adams JM, Faure H (2004) Review and Atlas of Palaeovegetation - Preliminary land ecosystem maps of the world since the Last Glacial Maximum. *Quaternary Environment Network* (http://www.esd.ornl.gov/projects/qen/adams4.html).

Amori G, Hutterer R, Kryštufek B, Yigit N, Mitsain G, Muñoz LJP, Aulagnier S (2008) *Jaculus jaculus*. *IUCN Red List of Threatened Species*, *Version 2009*, **2**. (Available at: http://www.iucnredlist.org).

An Z, Kutzbach JE, Prell WL, Porter SC (2001) Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since Late Miocene times. *Nature*, **411**, 62–66.

Ata AM, Shahin AAB (1999) Variation of G-bands in the chromosomes of Allactaga tetradactyla, *Jaculus jaculus jaculus* and *Jaculus orientalis* (Rodentia: Dipodidae) common in Egypt. *Journal of Union of Arab Biologists*, **11**, 295−309.

Ata AM, Shahin AAB (2006) C-heterochromatin and chiasma terminalization in the jerboas *Allactaga* and *Jaculus* (Rodentia: Dipodidae). *Belgian Journal of Zoology*, **136**, 59−67.

Ata AM, Shahin AAB, Allam HZ (2001) A comparative analysis of the rate of meiosis and chiasma terminalization in the jerboas *Allactaga* and *Jaculus* (Rodentia: Dipodidae) in Egypt. *Folia Biologica*, **49**,129−135.

Aulagnier S, Haffner P, Mitchell-Jones AJ, Moutou F, Zima J (2009) Mammals of Europe, North Africa and the Middle East. *A&C Black Publishers*.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Avise JC, Shapira JF, Daniel SW, Aquadro CF, Lansman RA (1983) Mitochondrial DNA differentiation during the speciation process in Peromyscus. *Molecular biology and evolution*, **1**, 38–56.

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the The Royal Society of London B: Biological sciences*, **265**, 1707–1712.

Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy*, **87**, 643–662.

Bannikova AA, Lebedev VS, Lissovsky AA,Matrosova V, Abramson NI, Obolenskaya EV, Tesakov AS (2009) Molecular phylogeny and evolution of the Asian lineage of

vole genus *Microtus* (Rodentia: Arvicolinae) inferred from mitochondrial cytochrome b sequence. *Biological Journal of the Linnean Society*, **99**, 595–613.

Begun DJ, Whitley P (2000) Reduced X-linked nucleotide polymorphism in Drosophila simulans. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 5960–5965.

Ben Faleh AR, Granjon L, Tatard C, Othmen AB, Said K, Cosson JF (2012a) Phylogeography of the Greater Egyptian Jerboa (*Jaculus orientalis*) (Rodentia: Dipodidae) in Mediterranean North Africa (J-N Volff, Ed,). *Journal of Zoology*, **286**, 208–220.

Ben Faleh AR, Granjon L, Tatard C, Boratyński Z, Cosson JF, Said K (2012b) Phylogeography of two cryptic species of African desert jerboas (Dipodidae: Jaculus). *Biological Journal of the Linnean Society*, **107**, 27–38.

Ben Faleh AR, Cosson JF, Tatard C, Othmen AB, Said K, Granjon L (2010a) Are there two cryptic species of the lesser Jerboa *jaculus jaculus* (Rodentia: Dipodidae) in tunisia? evidence from molecular, morphometric, and cytogenetic data. *Biological Journal of the Linnean Society*, **99**, 673–686.

Ben Faleh AR, Othmen AB, Said K (2010b) Taxonomy of the lesser jerboa *Jaculus* jaculus (Dipodidae, Rodentia) based on allozymic and morphological variation. *Current Zoology*, **56**, 421–431.

Bennett AW (1872) The Origin of Species by means of Natural Selection; or the Preservation of Favoured Races in the Struggle for Life. *Nature*, **5**, 318–319.

Besnard G, Rubio De Casas R, Vargas P (2007) Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea*). *Journal of Biogeography*, **34**, 736–752.

Bird CE, Fernandez-Silva I, Skillings DJ, Toonen RJ (2012) Sympatric Speciation in the Post “Modern Synthesis” Era of Evolutionary Biology. *Evolutionary Biology*, **39**, 158–180.

Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry RW, Huchon D (2009) Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evolutionary Biology*, **9**, 71.

Bonhomme M, Cuartero S, Blancher A, Crouau-Roy B (2009) Assessing natural introgression in 2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). *The Journal of heredity*, **100**, 158–69.

Boratyński Z, Brito JC, Campos JC, Karala M, Mappes T (2014a) Large spatial scale of the phenotype-environment color matching in two cryptic species of African desert jerboas (Dipodidae: *Jaculus*). *PLoS ONE*, **9**, e94342.

Boratyński Z, Brito JC, Mappes T (2012) The origin of two cryptic species of African desert jerboas (Dipodidae: *Jaculus*). *Biological Journal of the Linnean Society*, **105**, 435–445.

Boratyński Z, Melo-Ferreira J, Alves PC, Berto S, Koskela E, Pentikainen OT, Tarroso P, Ylilauri, Mappes T (2014b) Molecular and ecological signs of mitochondrial adaptation: consequences for introgression? *Heredity*, **113**, 1–10.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, **10**.

Bradley RD, Baker RJ (2001) A test of the genetic species concept: cytochrome-b sequences and mammals. *Journal of Mammalogy*, **82**, 960–973.

Brito JC, Godinho R, Martínez-Freiría F, Pleguezuelos JM, Rebelo H, Santos X, Vale CG, Velo-Antón G, Boratyński Z, Carvalho SB, Ferreira S, Gonçalves DV, Silva TL, Tarroso P, Campos JC, Leite JV, Nogueira J, Álvares F, Sillero N, Sow AS, Fahd S, Crochet P-A, Carranza S (2014) Unravelling biodiversity, evolution and threats to conservation in the sahara-sahel. *Biological Reviews*, **89**, 215–231.

Brouat C, Tatard C, Bâ K, Cosson J-F, Dobigny G, Fichet-Calvet E, Granjon L, Lecompte E, Loiseau A, Mouline K, Piry S, Duplantier J-M (2009) Phylogeography of the Guinea multimammate mouse (*Mastomys erythroleucus*): a case study for Sahelian species in West Africa. *Journal of Biogeography*, **36**, 2237–2250.

Butlin R, Debelle A, Kerth C, Snook RR, Beukeboom LW, Castillo Cajas RF, Diao W, Maan ME, Paolucci S, Weissing FJ, van de Zande L, Hoikkala A, Geuverink E, Jennings J, Kankare M, Knott KE, Tyukmaeva VI, Zoumadakis C, Ritchie MG, Barker D, Immonen E, Kirkpatrick M, Noor M, Macias Garcia C, Schmitt T, Schilthuizen M (2012) What do we need to know about speciation? *Trends in Ecology and Evolution*, **27**, 27–39.

Butlin RK, Galindo J, Grahame JW (2008) Review. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**, 2997–3007.

Carling MD, Brumfield RT (2008) Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in Passerina buntings. *Genetics*, **178**, 363–377.

Carneiro M, Ferrand N, Nachman MW (2009) Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, **181**, 593–606.

Carranza S, Arnold EN, Geniez P, Roca J, Mateo JA (2008) Radiation, multiple dispersal and parallelism in the skinks, Chalcides and Sphenops (Squamata: Scincidae), with comments on Scincus and Scincopus and the age of the Sahara Desert. *Molecular Phylogenetics and Evolution*, **46**, 1071–1094.

Carstens BC, Dewey TA (2010) Species delimitation using a combined coalescent and information-theoretic approach: An example from North American myotis bats. *Systematic Biology*, **59**, 400–414.

Carstens BC, Knowles LL (2007) Shifting distributions and speciation: Species divergence during rapid climate change. *Molecular Ecology*, **16**, 619–627.

Claussen M (2009) Late Quaternary vegetation-climate feedbacks. *Climate of the Past*, **5**, 203–216.

Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Colinvaux PA (1997) An arid Amazon? *Trends in ecology & evolution*, **12**, 318–319.

Corbet GB (1978) The mammals of the Palaearctic region: a taxonomic review. *Cornell University Press*.

Cosson J-F, Hutterer R, Libois R, Sarà M, Taberlet P, Vogel P (2005) Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia: Soricidae). *Molecular Ecology*, **14**, 1151–1162.

Cowling SA, Cox PM, Jones CD, Maslin MA, Peros M, Spall SA (2008) Simulated glacial and interglacial vegetation across Africa: Implications for species phylogenies and trans-African migration of plants and animals. *Global Change Biology*, **14**, 827–840.

Coyne JA, Orr HA (2004) Speciation. *Sinauer Associates*, 26-38.

Crandall KA, Bininda-emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Elsevier Science Ltd*, **15**, 290–295.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772–772.

Darwin C (1968) On the origin of species by means of natural selection. 1859. *Murray*.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, **292**, 673–679.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, **24**, 332–340.

deMenocal PB (2004) African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, **220**, 3–24.

De Queiroz K (2005) A Unified Concept of Species and Its Consequences for the Future of Taxonomy. *Proceedings of the California Academy of Sciences,* **56**, 196–215.

De Queiroz K (2007) Species concepts and species delimitation. *Systematic biology*, **56**, 879–886.

Douady CJ, Catzeflis F, Raman J, Springer MS, Stanhope MJ (2003) The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 8325–8330.

Drake NA, Blench RM, Armitage SJ, Bristow CS, White KH (2011) Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 458–462.

Drummond A, Rambaut A (2007) Tracer v1. 5. (Available from http://beast.bio.ed.ac.uk/Tracer).

Dumont HJ (1982) Relict distribution patterns of aquatic animals: another tool in in evaluating Late Pleistocene climate changes in the Sahara and Sahel. *Palaeoecology of Africa*, **14**, 1-24.

Dupont LM, Jahns S, Marret F, Ning S (2000) Vegetation change in equatorial West Africa: Time-slices for the last 150 ka. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **155**, 95–122.

Durant SM, Pettorelli N, Bashir S, Woodroffe R, Wacher T, De Ornellas P, Ransom C, Abáigar T, Abdelgadir M, El Alqamy H, Beddiaf M, Belbachir F, Belbachir-Bazi A, Berbash AA, Beudels-Jamar R, Boitani L, Breitenmoser C, Cano M, Chardonnet P, Collen B, Cornforth WA, Cuzin F, Gerngross P, Haddane B, Hadjeloum M, Jacobson A, Jebali A, Lamarque F, Mallon D, Minkowski K, Monfort S, Ndoassal B, Newby J, Ngakoutou BE, Niagate B, Purchase G, Samaïla S, Samna AK, Sillero-Zubiri C, Soultan AE, Stanley Price MR, Baillie JEM (2012) Forgotten Biodiversity in Desert Ecosystems. *Science*, **336**, 1379–1380.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

Edrey YH, Casper D, Huchon D, Mele J, Gelfond JA, Kristan DM, Nevo E, Buffenstein R (2012) Sustained high levels of neuregulin-1 in the longest-lived rodents; A key determinant of rodent longevity. *Aging Cell*, **11**, 213–222.

Edwards S V (2009) Is a new and general theory of molecular systematics emerging? *Evolution; international journal of organic evolution*, **63**, 1–19.

Endler JA (1977) Geographic variation, speciation, and clines. *Monographs in population biology*, **10**.

Estoup A, Garnery L, Solignac M, Cornuet JM (1995) Microsatellite variation in honey bee (*Apis mellifera* L.) populations: Hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, **140**, 679–695.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**, 2611–2620.

Flot JF (2010) Seqphase: A web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources*, **10**, 162–166.

Flot JF, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes*, **6**, 627–630.

Foley JA, Coe MT, Scheffer M, Wang G (2003) Regime Shifts in the Sahara and Sahel: Interactions between Ecological and Climatic Systems in Northern Africa. *Ecosystems*, **6**, 524–539.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schimitt J, Wilczek AM (2011) A Map of Local Adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–89.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Gaston KJ, Spicer JI (2004) Biodiversity: an Introduction. *Blackwell Science*, 15-15.

Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu H, Bulatova N, Ziv Y, Nachman MW (2008) Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*, **17**, 5349–5363.

Gharaibeh BM (1997) Systematics, distribution, and zoogeography of mammals of Tunisia. *Texas Tech University*.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of Southern European Refugia (ed Steven Weiss and Nuno Ferrand)*, 155–188.

Gonçalves DV, Brito JC, Crochet P-A, Geniez P, Padial JM, Harris DJ (2012) Phylogeny of north african agama lizards (reptilia: Agamidae) and the role of the sahara desert in vertebrate speciation. *Molecular Phylogenetics and Evolution*, **64**, 582–591.

Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *Journal of Heredity*, **86**, 485–486.

Granjon L, Duplantier JM, Catalan J, Brittondavidian J (1992) Karyotypic Data on Rodents from Senegal. *Israel Journal of Zoology*, **38**, 263–276.

Granjon L, Duplantier J (2009) Les rongeurs de l'Afrique sahélo-soudanienne. *Editions IRD, Publications Scientifiques du MNHN,* 82-84.

Griffin DL (2002) Aridity and humidity: Two aspects of the late Miocene climate of North Africa and the Mediterranean. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **182**, 65–91.

Guiller A, Coutellec-Vreto MA, Madec L, Deunff J (2001) Evolutionary history of the land snail *Helix aspersa* in the Western Mediterranean: preliminary results inferred from mitochondrial DNA sequences. *Molecular ecology*, **10**, 81–87.

Guiller A, Madec L (2010) Historical biogeography of the land snail *Cornu aspersum*: a new scenario inferred from haplotype distribution in the Western Mediterranean basin. *BMC evolutionary biology*, **10**, 18.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Haim A, Izhaki I (1995) Comparative physiology of thermoregulation in rodents: adaptations to arid and mesic environments. *Journal of Arid Environments*, **31**, 431–440.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. *Nucleic acids symposium series*, **41**, 95–98.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J (2011) Adaptation to Climate Across the Arabidopsis thaliana Genome. *Science*, **334**, 83–86.

Hare MP, Avise JC (1998) Population structure in the American oyster as inferred by nuclear gene genealogies. *Molecular biology and evolution*, **15**, 119–128.

Harrison DL (1978) A critical examination of alleged sibling species in the lesser three-toed jerboas (subgenus *Jaculus*) of the North African and Arabian deserts. *Bulletin of the Carnegie Museum of Natural History*, **6**, 77–80.

Harrison RG (1998) Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. *Endless Forms: Species and Speciation*, 19–31.

Haynes S, Jaarola M, Searle JB (2003) Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. *Molecular Ecology*, **12**, 951–956.

Heim De Balsac H (1936) Biogéographie des mammifères et des oiseaux de l'Afrique du Nord. *Bulletin biologique de la France et de la Belgique*.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC evolutionary biology*, **8**, 289.

Heled J, Drummond AJ (2010) Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution*, **27**, 570–580.

Hellborg L, Ellegren H (2003) Y chromosome conserved anchored tagged sequences (YCATS) for the analysis of mammalian male-specific DNA. *Molecular Ecology*, **12**, 283–291.

Hellborg L, Ellegren H (2004) Low Levels of Nucleotide Diversity in Mammalian Y Chromosomes. *Molecular Biology and Evolution*, **21**, 158–163.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role, in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hey J (2006) On the failure of modern species concepts. *Trends in Ecology and Evolution*, **21**, 447–450.

Hey J (2009) The Divergence of Chimpanzee Species and Subspecies as Revealed in Multipopulation Isolation-with-Migration Analyses. *Molecular Biology and Evolution*, **27**, 921–933.

Hey J (2010a) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hey J (2010b) Documentation for IMa2. *IMa2*, 1–60.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and *D. persimilis*. *Genetics*, **167**, 747–760.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**, 291–301.

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–485.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS genetics*, **6**, e1000862.

IUCN (2008) International Union for Conservation of Nature (http://www.iucnredlist.org/details/10912/0).

Le Houérou HN (1997) Climate, flora and fauna changes in the Sahara over the past 500 million years. *Journal of Arid Environments*, **37**, 619–647.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jaarola M, Martinkova N, Gunduz I, Brunhoffa C, Zimab J, Nadachowskie A, Amorif G, Bulatovag NS, Chondropoulosh B, Fraguedakis-Tsolish S, González-Estebani J, López-Fusterj MJ, Kandaurovk AS, Kefelioğlud H, Mathias ML, Villatei I, Searlem JB (2004) Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **33**, 647–663.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E, Searle S, Schmutz J, Grimwood J, Dickson MC, Myers RM, Miller CT, Summers BR, Knecht AK, Brady SD, Zhang H, Pollen AA, Howes T, Amemiya C, Baldwin J, Bloom T, Jaffe DB, Nicol R, Wilkinson J, Lander ES, Di Palma F, Lindblad-Toh K, Kingsley DM (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Jowers MJ, Amor F, Ortega P, Lenoir A, Boulay RR, Cerdá X, Galarza JA (2014) Recent speciation and secondary contact in endemic ants. *Molecular Ecology*, **23**, 2529–2542.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, **16**, 111–120.

Kirkpatrick M, Ravigné V (2002) Speciation by natural and sexual selection: models and experiments. *The American naturalist*, **159**, 22–35.

Knowles LL, Carstens BC (2007) Estimating a geographically explicit model of population divergence. *Evolution*, **61**, 477–493.

Kosman C, Breu H, Chappell G, Kumar S, Iasnopolski B, Kshirsargar B, Suri P, Paul R, Brown L, Mason G, Siu D, Sword E, Schoppe M, Pronyaev A, Bawge V (2001) Design and Performance Overview of SeqScape TM Software for Comparative Sequencing Analysis and Mutation Detection. *Human Genetics*, 8–8.

Kröpelin S, Verschuren D, Lézine A-M, Eggermont H, Cocquyt C, Francus P, Cazet J-P, Fagot M, Rumes B, Russell JM, Darius F, Conley DJ, Schuster M, von Suchodoletz H, Engstrom DR (2008) Climate-driven ecosystem succession in the Sahara: the past 6000 years. *Science*, **320**, 765–8.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Lebedev VS, Bannikova AA, Pagès M, Pisano J, Michaux JR, Shenbrot GI (2012) Molecular phylogeny and systematics of Dipodoidea: A test of morphology-based hypotheses. *Zoologica Scripta*, **42**, 231–249.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Maddison WP (1997) Gene Trees in Species Trees. *Systematic biology*, **46**, 523–536.

May RM (1994) Conceptual aspects of the quantification of the extent of biological diversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **345**, 13–20.

Mayden RL (1997) A hierarchy of species concepts: The denouement in the saga of the species problem. *Chapman and Hall*, 381-423.

Mayden RL (1999) Consilience and a hierarchy of species concepts: advances toward closure on the species puzzle. *Journal of nematology*, **31**, 95–116.

Mayr E (1942) Systematics and the origin of species, from the viewpoint of a zoologist. *Harvard University Press*, 102-150.

Mayr E (1963) Animal species and evolution. *Harvard University Press*, 1-30.

Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L, Alves PC (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology*, **61**, 367–381.

Millicent E, Thoday JM (1960) Gene flow and divergence under disruptive selection. *Science*, **131**, 1311–1312.

Montgelard C, Forty E, Arnal V, Matthee CA (2008) Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. *BMC evolutionary biology*, **8**, 321.

Moore WS, Aug N (1995) Inferring Phylogenies from Mtdna Variation - Mitochondrial-Gene Trees Versus Nuclear-Gene Trees. *Evolution*, **49**, 718–726.

Moritz C (1994) Defining Evolutionarily-Significant-Units for Conservation. *Trends in Ecology & Evolution*, **9**, 373–375.

Mouline K, Granjon L, Galan M, Tatard C, Abdoullaye D, Agatteyine S, Duplantier J-M, Cosson J-F (2008) Phylogeography of a Sahelian rodent species *Mastomys huberti*: A Plio-Pleistocene story of emergence and colonization of humid habitats. *Molecular Ecology*, **17**, 1036–1053.

Moutinho F, Qninba A, Harrington A, Forbes K, Mediani M, Sérén N, Mappes T, Boratyński Z (2015) Winter breeding of the Lesser Egyptian Jerboa *Jaculus jaculus* (Linnaeus,1758) in Southern Morocco. *Go-South Bull*, **12** , 24–27.

Nabholz B, Glémin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals - The longevity hypothesis. *Molecular Biology and Evolution*, **25**, 120–130.

Naomi SI (2011) On the integrated frameworks of species concepts: Mayden's hierarchy of species concepts and de Queiroz's unified concept of species. *Journal of Zoological Systematics and Evolutionary Research*, **49**, 177–184.

Narum S, Buerkle C, Davey J, Miller M, Hohenlohe P (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular ecology*, **22**, 2841–2847.

Neafsey DE, Lawniczak MKN, Park DJ, Redmond SN, Coulibaly MB, Traoré SF, Sagnon N, Costantini C, Johnson C, Wiegand RC, Collins FH, Lander ES, Wirth DF, Kafatos FC, Besansky NJ, Christophides GK, Muskavitch MAT (2010) SNP

genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science,* **330**, 514–517.

Nicolas V, Granjon L, Duplantier JM, Cruaud C, Dobigny G (2009) Phylogeography of spiny mice (genus *Acomys*, Rodentia: Muridae) from the south-western margin of the sahara with taxonomic implications. *Biological Journal of the Linnean Society*, **98**, 29–46.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, D'amico JA, Itoua I, Strand HE, Morrison JC, Loucks CJ, Allnutt TF, Ricketts TH, Kura Y, Lamoreux JF, Wettengel WW, Hedao P, Kassem KR (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, **51**, 933-938.

Osborn DJ, Helmy I (1980) The contemporary land mammals of Egypt including Sinai. *Fieldiana Zoology*, **5**.

Palo JU, Schmeller DS, Laurila A, Primmer CR, Kuzmin SL, Merila J (2004) High degree of population subdivision in a widespread amphibian. *Molecular Ecology*, **13**, 2631–2644.

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular biology and evolution*, **11**, 426–435.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular biology and evolution*, **5**, 568–583.

Parmesan C, Yohe G, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Pauls SU, Nowak C, Bálint M, Pfenninger M (2013) The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, **22**, 925–946.

Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB (2012) Cryptic speciation in the field vole: A multilocus approach confirms three highly divergent lineages in Eurasia. *Molecular Ecology*, **21**, 6015–6032.

Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pellegrino KCM, Rodrigues MT, Waite AN, Morando M, Yassuda YY, Sites JR JW (2005) Phylogeography and species limits in the *Gymnodactylus darwinii* complex

(Gekkonidae, Squamata): Genetic structure coincides with river systems in the Brazilian Atlantic Forest. *Biological Journal of the Linnean Society*, **85**, 13–26.

Peter F (1961) Répartition géographique et écologique des rongeurs désertiques. *Mammalia*, **25**.

Pinho C, Hey J (2010) Divergence with Gene Flow: Models and Data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215–230.

Pisano J, Condamine FL, Lebedev V, Bannikova A, Quéré J-P, Shenbrot GI, Pagès M, Michaux JR (2015) Out of Himalaya: the impact of past Asian environmental changes on the evolutionary and biogeographical history of Dipodoidea (Rodentia). *Journal of Biogeography*, **42**, 856-870.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Procaccini G, Olsen JL, Reusch TBH (2007) Contribution of genetics and genomics to seagrass biology and conservation. *Journal of Experimental Marine Biology and Ecology*, **350**, 234–259.

Rambaut A (2009) FigTree v1.3.1. (Program package available at http://tree.bio.ed.ac).

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular biology and evolution*, **19**, 2092–2100.

Ranck GL (1968) The rodents of Libya: taxonomy, ecology and zoogeographical relationship. *Bulletin of Smithsonian Institution Museum of Natural History*, **275**.

Rato CBJC, Carretero MA, Shacham SLB, Harris DJ (2007) Phylogeography and genetic diversity of *Psammophis schokari* (Serpentes) in North Africa based on mitochondrial DNA sequences. *Journal of African zoology*, **42**, 112–117.

Raymo M (1994) The Initiation of Northern Hemisphere Glaciation. *Annual Review of Earth and Planetary Sciences*, **22**, 353–383.

Rice WR (1987) Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. *Evolutionary Ecology*, **1**, 301–314.

Riddle BR, Dawson MN, Hadly EA, Hafner DJ, Hickerson MJ, Mantooth SJ, Yoder AD (2008) The role of molecular genetics in sculpting the future of integrative biogeography. *Progress in Physical Geography*, **32**, 173–202.

Rognon P (1993) Biogéographie d'un désert. *L'Harmattan, Paris*.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) Mrbayes 3.2: Efficient bayesian

phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Rosen DE (1978) Vicariant Patterns and Historical Explanation in Biogeography. *Systematic Zoology*, **27**, 159-188.

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature reviews. Genetics*, **3**, 380–390.

Ruddiman WF, Sarnthein M, Backman J, Baldauf JG, Curry W, Dupont LM, Janecek T, Pokras EM, Raymo ME, Stabell B, Stein R, Tiedemann R (1989) Late Miocene to Pleistocene evolution of climate in Africa and low-latitude Atlantic: Overview of Leg 108 results. *Proceeding of the Ocean Drilling Program, Scientific Results,* **108**, 463–484.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Santos AM, Cabezas MP, Tavares AI, Xavier R, Branco M (2015) tcsBU: a tool to extend TCS network layout and visualization. *Bioinformatics*, **btv636**.

Sax DF, Stachowicz JJ, Brown JH, Bruno JF, Dawson MN, Gaines SD, Grosberg RK, Hastings A, Holt RD, Mayfield MM, O'Connor MI, Rice WR (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology and Evolution*, **22**, 465–471.

Sayre R, Comer P, Hak J, Josse C, Bow J, Warner H, Larwanou M, Kelbessa E, Bekele T, Kehl H, Amena R, Andriamasimanana R, Ba T, Benson T, Boucher T, Brown M, Cress J, Dassering O, Friesen B, Gachathi F, Houcine S, Keita M, Khamala E, Marangu D, Mokua F, Morou B, Mucina L, Mugisha S, Mwavu E, Rutherford M, Sanou P, Syampungani S, Tomor B, Vall A, Vande Weghe J, Wangui E, Waruingi L (2013) A new map of standardized terrestrial ecosystems of Africa. *Washington, DC: Association of American Geographers*, 1–24.

Schaffner SF (2004) The X chromosome in population genetics. *Nature Reviews Genetics*, **5**, 43–51.

Schilthuizen M (2000) Dualism and conflicts in understanding speciation. *BioEssays*, **22**, 1134–1141.

Schlötterer C (2004) The evolution of molecular markers--just a matter of fashion? *Nature reviews. Genetics*, **5**, 63–69.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372–380.

Schuster M, Duringer P, Ghienne J-F, Vignaud P, Mackaye HT, Likius A, Brunet M (2006) The age of the Sahara desert. *Science*, **311**, 821.

Scoble J, Lowe AJ (2010) A case for incorporating phylogeography and landscape genetics into species distribution modelling approaches to improve climate adaptation and conservation planning. *Diversity and Distributions*, **16**, 343–353.

Seifert B (2009) Cryptic species in ants (Hymenoptera: Formicidae) revisited: we need a change in the alpha-taxonomic approach. *Myrmecological News*, **12**, 149–166.

Sequeira F, Alexandrino J, Weiss S, Ferrand N (2008) Documenting the advantages and limitations of different classes of molecular markers in a well-established phylogeographic context: Lessons from the Iberian endemic Golden-striped salamander, *Chioglossa lusitanica* (Caudata: Salamandridae). *Biological Journal of the Linnean Society*, **95**, 371–387.

Shahin AAB (1999) A comparative study of the molar and soft palate characters of the genera *Allactaga* and *Jaculus* (Mammalia: Rodentia) in Egypt. *Zoology in the Middle East*, **18**, 17–32.

Shahin AAB (2003) Genetic differentiation and relationship of the dipodids *Allactaga* and *Jaculus* (Mammalia,Rodentia) in Egypt based on protein variation. *Acta Theriologica*, **48**, 309–324.

Shahin AAB (2005) Growth and maturation of metatarsals and their taxonomic significance in the jerboas *Allactaga* and *Jaculus* (Rodentia: Dipodidae*). Acta Zoologica*, **86**, 81−90.

Shahin AAB, Ata AM (2001) A comparative study on the Karyotype and meiosis of the jerboas *Allactaga* and *Jaculus* (Rodentia: Dipodidae) in Egypt. *Zoology in the Middle East*, **22**, 5–16.

Shahin AAB, Ata AM (2004) C-banding karyotype and relationship of the dipodids *Allactaga* and *Jaculus* (Mammalia: Rodentia) in Egypt. *Folia Biologica*, **52**, 25−31.

Shenbrot GI, Krasnov B, Rogovin KA (2012) Spatial ecology of desert rodent communities. *Springer Science & Business Media*, 151-193.

Shenbrot GI, Sokolov VE, Heptner VG, Koval'skaya YM (2008) Mammals of Russia and adjacent regions: jerboas. *Amerind Publishing Co. Pvt. Ltd*.

Simpson GG (1951) The species concept. *Evolution*, **5**, 285–298.

Sites JW, Marshall JC (2003) Delimiting species: A Renaissance issue in systematic biology. *Trends in Ecology and Evolution*, **18**, 462–470.

Smith JM (1966) Sympatric speciation. *American Naturalist*, **100**, 637–650.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American journal of human genetics*, **76**, 449–462.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American journal of human genetics*, **68**, 978–989.

Stevison LS, Kohn MH (2009) Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Molecular Ecology*, **18**, 2457–2475.

Strasburg JL, Rieseberg LH (2010) How Robust Are "Isolation with Migration" Analyses to Violations of the IM Model? A Simulation Study. *Molecular Biology and Evolution*, **27**, 297–310.

Swezey CS (2009) Cenozoic stratigraphy of the Sahara, Northern Africa. *Journal of African Earth Sciences*, **53**, 89–121.

Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, **68**, 41–55.

Taberlet P, Luikart G, Waits LP (1999) Noninvasive genetic sampling: Look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Templeton AR (1981) Mechanisms of speciation - a population genetic approach. *Annual Review of Ecology and Systematics*, **12**, 23–48.

Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology*, **10**, 779–791.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, **22**, 4673–4680.

Triant DA, DeWoody JA (2006) Accelerated molecular evolution in *Microtus* (Rodentia) as assessed via complete mitochondrial genome sequences. *Genetica*, **128**, 95–108.

UNDP (2010) United Nations Development Programme. (http://hdr.undp.org/en/reports/global/hdr2010/).

Vallone PM, Butler JM (2004) AutoDimer: A screening tool for primer-dimer and hairpin structures. *BioTechniques*, **37**, 226–231.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Van Zinderen BEM (1978) Quaternary vegetation changes in southern Africa. *Biogeography and ecology of Southern Africa*, **1**, 131–143.

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 451–460.

Wagner P, Melville J, Wilms TM, Schmitz A (2011) Opening a box of cryptic taxa - the first review of the North African desert lizards in the *Trapelus mutabilis* (Merre 1820) complex (Squamata: Agamidae) with descriptions of new taxa. *Zoological Journal of the Linnean Society*, **163**, 884–912.

Wakeley J (1994) The variance of pairwise nucleotide differences in two populations with migration. *Theoretical population biology*, **49**, 39–57.

Wan QH, Wu H, Fujihara T, Fang SG (2004) Which genetic marker for which conservation genetics issue? *Electrophoresis*, **25**, 2165–2176.

Wang Y, Notaro M, Liu Z, Gallimore R, Levis S, Kutzbach JE (2008) Detecting vegetation-precipitation feedbacks in mid-Holocene North Africa from two climate models. *Climate of the Past, European Geosciences Union (EGU)*, **4**, 59-67.

Wang J, Wang YJ, Liu ZC, Li JQ, Xi P (1999) Cenozoic environmental evolution of the Qaidam Basin and its implications for the uplift of the Tibetan Plateau and the drying of central Asia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **152**, 37–47.

Ward D (2009) The biology of deserts. *Oxford University Press*.

Weiss S, Ferrand N (2007) Current perspectives in phylogeography and the significance of South European refugia in the creation and maintenance of European biodiversity. *Phylogeography of Southern European Refugia*, 341 – 357.

Wiley EO (1978) The Evolutionary Species Concept. *Systematic Zoology*, **27**, 17–26.

Wilson DE, Reeder DM (2005) Mammal Species of the World: a taxonomic and geographic reference. *Jonhs Hopkins University Press*, **1**, 882-884.

Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–1853.

Won Y-J, Hey J (2005) Divergence Population Genetics of Chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.

Wright S (1940) The statistical consequences of Mendelian heredity in relation to speciation. *The new systematics*, 161-183.

Wu CI (1991) Inferences of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics*, **127**, 429–435.

Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular ecology*, **12**, 563–584.

Zhang Q, Xia L, Kimura Y, Shenbrot G, Zhang Z, Ge D, Yang Q (2012) Tracing the Origin and Diversification of Dipodoidea (Order: Rodentia): Evidence from Fossil Record and Molecular Phylogeny. *Evolutionary Biology*, **40**, 32–44.

Zheng H (2004) Late Miocene and mid-Pliocene enhancement of the East Asian monsoon as viewed from the land and sea. *Global and Planetary Change*, **41**, 147–155.

Annexes

Annex 1. List of the complete *J. jaculus* samples used in analysis. The two columns labelled as "*Cytb* (short/long fragment)" and "Nuclear markers" indicate the samples that successfully amplified for at least one of the markers. The last two columns show the GenBank accession numbers for the samples that were analysed with *Cytb* and *UWF* genes in previous studies, but were analysed again in this survey. The respective mitochondrial clade is displayed. *J. orientalis* samples used are also included. The samples geographic distribution is shown in Figure 3; although for some samples the exact coordinates are not available.

Sample Code	Museum/Field Collections	Country	Lat	Long	Clade	*Cytb* (short/long fragment)	Nuclear Markers	Cytb GB	UWF GB
ZBSC 0013	Mauritania 2010	Morocco	28,829	-10,266	2				
ZBSC 0019	Mauritania 2010	Mauritania	20,997	-16,283	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214535[a]	
ZBSC 0020	Mauritania 2010	Mauritania	20,929	-16,221	2	Short	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0021	Mauritania 2010	Mauritania	20,602	-16,012	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214536[a]	JN227872[a]
ZBSC 0027	Mauritania 2010	Mauritania	16,566	-14,198	2				JN227875[a]
ZBSC 0028	Mauritania 2010	Mauritania	16,435	-14,037	2				JN227876[a]
ZBSC 0064	Mauritania 2010	Mauritania	18,901	-15,416	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214539[a]	JN227875[a]
ZBSC 0070	Mauritania 2010	Mauritania	20,724	-16,057	2	Short/Long	*ADRA2B, DBX5, M*	JN214540[a]	
ZBSC 0072	Mauritania 2010	Mauritania	20,379	-15,991	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214541[a]	JN227876[a]
ZBSC 0079	Mauritania 2010	Mauritania	20,613	-16,013	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0081	Mauritania 2010	Western Sahara (Morocco)	22,639	-16,337	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214542[a]	
ZBSC 0082	Mauritania 2010	Western Sahara (Morocco)	24,297	-15,333	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214543[a]	
ZBSC 0083	Mauritania 2010	Western Sahara (Morocco)	24,630	-14,945	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214544[a]	
ZBSC 0084	Mauritania 2010	Western Sahara(Morocco)	25,322	-14,795	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214545[a]	JN227877[a]
ZBSC 0193	Mauritania 2011	Western Sahara (Morocco)	25,267	-14,821	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0196	Mauritania 2011	Western Sahara	24,006	-15,611	2	Short/Long	*ADRA2B, DBX5, M*		

		(Morocco)							
ZBSC 0197	Mauritania 2011	Western Sahara (Morocco)	22,829	-16,250	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663500[b]	
ZBSC 0198	Mauritania 2011	Western Sahara (Morocco)	22,557	-16,370	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663501[b]	
ZBSC 0218	Mauritania 2011	Mauritania	21,438	-12,980	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663502[b]	
ZBSC 0219	Mauritania 2011	Mauritania	21,352	-13,039	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663503[b]	
ZBSC 0224	Mauritania 2011	Mauritania	20,557	-12,572	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663582[b]	
ZBSC 0226	Mauritania 2011	Mauritania	20,508	-12,831	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663504[b]	
ZBSC 0240	Mauritania 2011	Mauritania	20,254	-13,296	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663505[b]	
ZBSC 0241	Mauritania 2011	Mauritania	20,253	-13,311	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663506[b]	
ZBSC 0242	Mauritania 2011	Mauritania	20,016	-13,887	2	Short/Long	*ADRA2B, GHR, IRBP, DBX5, M*	KC663583[b]	
ZBSC 0243	Mauritania 2011	Mauritania	19,651	-14,504	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663507[b]	
ZBSC 0244	Mauritania 2011	Mauritania	19,651	-14,504	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663508[b]	
ZBSC 0245	Mauritania 2011	Mauritania	19,651	-14,504	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663509[b]	
ZBSC 0246	Mauritania 2011	Mauritania	19,651	-14,504	1	Short	*GHR*		
ZBSC 0256	Mauritania 2011	Mauritania	17,591	-12,848	1	Short/Long	*ADRA2B, DBX5, M*	KC663581[b]	
ZBSC 0265	Mauritania 2011	Mauritania	18,094	-12,132	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	KC663510[b]	
ZBSC 0267	Mauritania 2011	Western Sahara (Morocco)	22,136	-16,570	2	Short/Long	*GHR, DBX5*	KC663511[b]	
ZBSC 0290	Mauritania 2012	Western Sahara (Morocco)	27,149	-10,847					
ZBSC 0291	Mauritania 2012	Western Sahara (Morocco)	27,055	-11,410	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0292	Mauritania 2012	Western Sahara (Morocco)	26,959	-11,657	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0293	Mauritania 2012	Western Sahara (Morocco)	26,933	-11,700	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0294	Mauritania 2012	Western Sahara (Morocco)	26,811	-11,746	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0295	Mauritania 2012	Western Sahara (Morocco)	25,246	-12,488	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0296	Mauritania 2012	Western Sahara (Morocco)	27,167	-10,964	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		

ZBSC 0303	Mauritania 2012	Mauritania	21,021	-16,304	1	Short	*GHR, IRBP, UWF, DBX5, M*		
ZBSC 0306	Mauritania 2012	Mauritania	16,633	-15,196	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0382	Mauritania: October/December 2012	Morocco	32,474	-4,494					
ZBSC 0383	Mauritania: October/December 2012	Morocco	28,961	-10,508	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0384	Mauritania: October/December 2012	Morocco	28,394	-11,026	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0385	Morocco 2012	Morocco	28,330	-10,913	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0388	Morocco 2012	Morocco	28,823	-10,371	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0403	Mauritania 2014	Western Sahara (Morocco)	25,936	-14,514	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0404	Mauritania 2014	Western Sahara (Morocco)	25,654	-14,660	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0408	Mauritania 2014	Western Sahara (Morocco)	23,846	-15,863	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0413	Mauritania 2014	Western Sahara (Morocco)	23,595	-15,715	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0414	Mauritania 2014	Western Sahara (Morocco)	23,115	-14,964	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0417	Mauritania 2014	Mauritania	20,844	-16,149	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0420	Mauritania 2014	Mauritania	20,093	-15,927	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0423	Mauritania 2014	Mauritania	16,220	-13,260	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0424	Mauritania 2014	Mauritania	15,565	-12,327	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0509	Mauritania 2014	Mauritania	18,383	-9,313	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0526	Mauritania 2014	Mauritania	18,559	-11,248	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0542	Mauritania 2014	Mauritania	18,357	-11,816	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0543	Mauritania 2014	Mauritania	18,357	-11,816	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0558	Mauritania 2014	Mauritania	18,109	-11,916	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0559	Mauritania 2014	Mauritania	18,000	-11,884	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0591	Mauritania 2014	Morocco	32,255	-2,215	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		

ZBSC 0599	Mauritania 2014	Morocco	32,156	-1,327	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0614	Mauritania 2014	Morocco	28,393	-11,028	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0626	Mauritania 2014	Morocco	28,179	-11,857	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0629	Mauritania 2014	Morocco	28,559	-10,918	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0647	Morocco winter 2015	Morocco	27,924	-11,444	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0650	Morocco winter 2015	Morocco	27,907	-11,600	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0651	Morocco winter 2015	Morocco	27,895	-11,610	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0652	Morocco winter 2015	Morocco	27,901	-11,605	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0653	Morocco winter 2015	Morocco	27,875	-11,588	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0654	Morocco winter 2015	Morocco	27,907	-11,600	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0665	Morocco winter 2015	Morocco	27,224	-12,883	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0675	Morocco winter 2015	Morocco	26,579	-12,769	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0678	Morocco winter 2015	Morocco	27,046	-11,537	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0679	Morocco winter 2015	Morocco	27,054	-11,432	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0681	Morocco winter 2015	Morocco	27,076	-11,756	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0683	Morocco winter 2015	Morocco	27,386	-11,694	2	Short/Long			
ZBSC 0684	Morocco winter 2015	Morocco	27,453	-11,681	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0685	Morocco winter 2015	Morocco	27,453	-11,681	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0686	Morocco winter 2015	Morocco	27,575	-11,634	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0688	Morocco winter 2015	Morocco	27,938	-11,577	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0690	Morocco winter 2015	Morocco	27,902	-11,605	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0691	Morocco winter 2015	Morocco	27,918	-11,553	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0692	Morocco winter 2015	Morocco	27,918	-11,554	2	Short/Long	*ADRA2B, GHR, IRBP, UWF,*		

							DBX5, M		
ZBSC 0693	Morocco winter 2015	Morocco	27,914	-11,545	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0694	Morocco winter 2015	Morocco	27,902	-11,604	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0695	Morocco winter 2015	Morocco	27,927	-11,571	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0696	Morocco winter 2015	Morocco	27,907	-11,601	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0697	Morocco winter 2015	Morocco	27,938	-11,577	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0699	Morocco winter 2015	Morocco	27,926	-11,444	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
ZBSC 0703	Morocco winter 2015	Morocco	28,110	-11,301	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
D100	National Geographic:Sept-Decem 2004	Mauritania	17,938	-12,267	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214504[a]	JN214554[a]
D1283	National Geographic:Sept-Decem 2004	Western Sahara (Morocco)	22,446	-16,448	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214509[a]	JN214552[a]
D144	National Geographic:Sept-Decem 2004	Western Sahara (Morocco)	24,847	-14,844	2	Short/Long	*ADRA2B, GHR, IRBP, DBX5, M*	JN214510[a]	
D145	National Geographic:Sept-Decem 2004	Western Sahara (Morocco)	25,245	-14,821	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214511[a]	
D535	National Geographic:Sept-Decem 2004	Western Sahara (Morocco)	21,937	-16,875	2	Short/Long	*ADRA2B, UWF*	JN214521[a]	JN214553[a]
D549	National Geographic:Sept-Decem 2004	Western Sahara (Morocco)	24,788	-14,865	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214523[a]	
D576	National Geographic:Sept-Decem 2004	Morocco	29,389	-8,129	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214524[a]	JN214549[a]
D577	National Geographic:Sept-Decem 2004	Morocco	30,038	-6,894	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214525[a]	JN214548[a]
D578	National Geographic:Sept-Decem 2004	Morocco	30,038	-6,894	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		JN227870[a]
D6	National Geographic:Sept-Decem 2004	Morocco	28,469	-11,021	2			JN214526[a]	
D796	National Geographic:Sept-Decem 2004	Western Sahara (Morocco)	24,686	-14,862	2	Short/Long	*ADRA2B, GHR, UWF, DBX5*	JN214528[a]	JN214551[a]
D800	National Geographic:Sept-Decem 2004	Western Sahara (Morocco)	25,306	-14,803	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, M*	JN214529[a]	
D684	National Geographic:Sept-Decem 2004	Morocco	31,075	-4,011	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5*	JN214533[a]	JN214547[a]
D53	National Geographic:Sept-Decem 2004	Mauritania	17,195	-7,141	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214531[a]	JN214558[a]
D52	National Geographic:Sept-Decem 2004	Mauritania	17,225	-7,069	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214520[a]	

D117	National Geographic:Sept-Decem 2004	Mauritania	17,393	-13,453	1	Short/Long	*GHR, DBX5*	JN214508[a]	
D493	National Geographic:Sept-Decem 2004	Mauritania	17,408	-16,062	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214517[a]	JN214559[a]
D113	National Geographic:Sept-Decem 2004	Mauritania	17,693	-12,571	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214507[a]	
D3055	National Geographic:Sept-Decem 2004	Mauritania	17,895	-11,716	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214514[a]	JN214561[a]
D22	National Geographic:Sept-Decem 2004	Mauritania	17,899	-12,334	1	Short/Long	*ADRA2B, GHR, UWF, DBX5, M*	JN214513[a]	
D101	National Geographic:Sept-Decem 2004	Mauritania	17,938	-12,267	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214506[a]	
D3107	National Geographic:Sept-Decem 2004	Mauritania	18,021	-12,050	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214515[a]	JN214556[a]
D506	National Geographic:Sept-Decem 2004	Mauritania	19,439	-14,754	1	Short/Long	*ADRA2B, GHR, IRBP, UWF*	JN214518[a]	JN214557[a]
D511	National Geographic:Sept-Decem 2004	Mauritania	19,641	-14,522	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214519[a]	
D1003	National Geographic:Sept-Decem 2004	Mauritania	20,378	-15,991	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214505[a]	
D1630	National Geographic:Sept-Decem 2004	Mauritania	21,355	-13,025	1				
D945	National Geographic:Sept-Decem 2004	Morocco	28,633	-10,753	1				
D320	National Geographic:Sept-Decem 2004	Tunisia	33,014	10,952	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214516[a]	JN214560[a]
D536	National Geographic:Sept-Decem 2004	Western Sahara (Morocco)	21,969	-16,874	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214532[a]	JN214555[a]
D541	National Geographic:Sept-Decem 2004	Western Sahara (Morocco)	22,367	-16,462	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*	JN214522[a]	
D316	National Geographic:Sept-Decem 2004	Tunisia	33,498	9,383	2		*ADRA2B, GHR, UWF*		
8067	Collected by Luis Garcia-Cardenete	Morocco	28,991	-10,315	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
8069	Collected by Luis Garcia-Cardenete	Morocco	28,598	-10,859	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
8070	Collected by Luis Garcia-Cardenete	Morocco	28,298	-11,189	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
8072	Collected by Luis Garcia-Cardenete	Morocco	27,723	-11,565	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
8076	Collected by Luis Garcia-Cardenete	Western Sahara (Morocco)	26,689	-11,854	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
8345	Collected by Luis Garcia-Cardenete	Morocco	32,416	-2,087	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
8350	Collected by Luis Garcia-Cardenete	Morocco	32,141	-2,791	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
8367	Collected by Luis Garcia-Cardenete	Morocco	29,825	-7,524	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
8374	Collected by Luis Garcia-Cardenete	Morocco	28,905	-10,170	2	Short/Long	*ADRA2B, GHR, IRBP, UWF,*		

							DBX5, M		
8545	Observed by F. Martínez-Freiría	Morocco	31,144	-7,404		Short	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
9072	Collected by LG Cardenete	Morocco	28,069	-11,361	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
9083	Collected by LG Cardenete	Morocco	28,361	-10,859	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
9086	Collected by LG Cardenete	Morocco	29,010	-10,207	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
9087	Collected by LG Cardenete	Morocco	28,655	-10,486	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
10355	F Álvares, R Godinho, M Nakamura, J Layna	Senegal	16,355	-16,224	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
6064	Mauritania 2011	Mauritania	18,236	-11,519	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
6269	Mauritania 2011	Mauritania	18,480	-16,022	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
6367	Mauritania 2011	Mauritania	19,960	-16,084	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
6267	Mauritania 2011	Mauritania	18,480	-16,022	1				
6268	Mauritania 2011	Mauritania	18,480	-16,022	1	Short	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
6481	Mauritania 2011	Western Sahara (Morocco)	22,276	-16,494	2		*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
6482	Mauritania 2011	Western Sahara (Morocco)	22,348	-16,469	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
1221	Mauritania 2011	Morocco	31,075	-4,011					
11043	Mauritania: November 2014	Mauritania	16,068	-11,509	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
11379	Mauritania: November 2014	Mauritania	18,197	-15,047	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
11382	Mauritania: November 2014	Mauritania	18,261	-14,981	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
11384	Mauritania: November 2014	Mauritania	18,490	-14,644	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
11408	Mauritania: November 2014	Mauritania	18,429	-14,801	1	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
11411	Mauritania: November 2014	Western Sahara (Morocco)	22,061	-16,736	2	Short/Long	*ADRA2B, GHR, IRBP, UWF, DBX5, M*		
37709	Royal Museum of Central Africa Tervuren	Egypt	22,004	25,145	1	Short			
37711	Royal Museum of Central Africa Tervuren	Egypt	22,004	25,145					
37720	Royal Museum of Central Africa Tervuren	Egypt	22,004	25,145	1	Short			
37728	Royal Museum of Central Africa Tervuren	Egypt	22,004	25,145	1	Short/Long			

37731	Royal Museum of Central Africa Tervuren	Egypt	22,004	25,145	1	Short			
37750	Royal Museum of Central Africa Tervuren	Egypt	22,030	25,141	1	Short/Long			
37761	Royal Museum of Central Africa Tervuren	Egypt	22,037	25,097	1	Short			
37762	Royal Museum of Central Africa Tervuren	Egypt	22,037	25,097	1	Short			
37763	Royal Museum of Central Africa Tervuren	Egypt	22,037	25,097	1	Short			
37764	Royal Museum of Central Africa Tervuren	Egypt	22,004	25,145	1	Short			
5046	Royal Museum of Central Africa Tervuren	Sudan	15,600	32,800	1	Short			
5047	Royal Museum of Central Africa Tervuren	Sudan	15,600	32,800	1	Short			
13536	Natural History Museum (Naturhistorisches Museum) – Vienna	Israel			1	Short/Long			
16984	Natural History Museum, Brussels, Belgium	Egypt	22,010	25,140	1	Short			
16989	Natural History Museum, Brussels, Belgium	Egypt	22,010	25,140	1	Short/Long			
16992	Natural History Museum, Brussels, Belgium	Egypt	22,037	25,096	1	Short			
16993	Natural History Museum, Brussels, Belgium	Egypt	22,037	25,096	1	Short			
467B	Natural History Museum, Brussels, Belgium	Saudi-Arabia			1	Short			
72.64.1.	Hungarian Museum of natural History	Tunisia	36,480	10,670					
72.64.2.	Hungarian Museum of natural History	Tunisia	36,480	10,670					
72.64.3.	Hungarian Museum of natural History	Tunisia	36,480	10,670					
72.64.4.	Hungarian Museum of natural History	Tunisia	36,480	10,670					
72.64.5.	Hungarian Museum of natural History	Tunisia	36,480	10,670					
81.54.1	HNHM, Hungary	Iraq							
86.28.1.	Hungarian Museum of natural History	Algeria	32,833	3,767					
88.119.1.	Hungarian Museum of natural History	Tunisia	36,480	10,670					
2000.75.7	HNHM, Hungary	Iraq							
3302	MZUF, Florence	Somalia	4,720	46,597					
6262	MZUF, Florence	Somalia	8,237	48,265					
3299	MZUF, Florence	Somalia	4,720	46,597					
6261	MZUF, Florence	Somalia	4,720	46,597					
6296	MZUF, Florence	Somalia	9,975	50,540					
10010	MZUF, Florence	Somalia	4,083	46,550					
10011	MZUF, Florence	Somalia	4,083	46,550					
2484	MZUF, Florence	Somalia	4,720	46,597					
6260	MZUF, Florence	Somalia	4,720	46,597					
M/9798/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of	Algeria	28,250	0,200	1	Short			

	Mammals								
M/9799/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of Mammals	Algeria	30,717	3,133	1	Short			
M/9796/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of Mammals	Algeria	33,100	1,267					
M/9795/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of Mammals	Algeria	32,166	5,189					
M/9801/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of Mammals	Algeria	35,137	3,016	1	Short			
M/9800/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of Mammals	Algeria	23,700	5,133					
M/9797/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of Mammals	Algeria	28,250	-0,200	1	Short			
M/9794/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of Mammals	Algeria	35,094	3,017					
M/9790/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of Mammals	Algeria							
M/9792/90	Polish Academy of Science, Museum of Institute of Systematics and Evolution of Mammals	Algeria							
5606	MZUF Florence	Libya							
22020	Belgium Royal Museum of Natural History Brussels	Tunisia			1	Short			
22021	Belgium Royal Museum of Natural History Brussels	Tunisia	34,483	9,583	1	Short			
30263	Natural History Museum (Naturhistorisches Museum) – Vienna	Libya			1	Short			
30264	Natural History Museum (Naturhistorisches Museum) – Vienna	Libya							
M-0274	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia							
M-0275	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia			1	Short			
M-0277	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia							
M-0278	Belgium Royal Museum of Cantral Africs	Tunisia							

	Tervuren								
M-0279	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia							
M-0280	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia							
M-0281	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia			1	Short			
M-0282	Belgium Royal Museum of Cantral Africs Tervuren	Libya			1	Short			
M-0283	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia			2	Short			
M-0284	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia							
M-0285	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia			2	Short			
M-0286	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia							
M-0287	Belgium Royal Museum of Cantral Africs Tervuren	Tunisia			2	Short			
11908	Natural History Museum (Naturhistorisches Museum) – Vienna	Tunisia							
9812	Belgium Royal Museum of Natural History Brussels	Libya			2	Short			
9813	Belgium Royal Museum of Natural History Brussels	Libya							
9814	Belgium Royal Museum of Natural History Brussels	Libya							
9815	Belgium Royal Museum of Natural History Brussels	Libya							
9816	Belgium Royal Museum of Natural History Brussels	Libya							
9817	Belgium Royal Museum of Natural History Brussels	Libya							
9818	Belgium Royal Museum of Natural History Brussels	Libya							
18044	Belgium Royal Museum of Natural History Brussels	Libya	30,766	17,783	1				
18043	Belgium Royal Museum of Natural History Brussels	Libya	30,833	17,783	1				
18042	Belgium Royal Museum of Natural History Brussels	Libya	30,833	17,783	2	Short			
18041	Belgium Royal Museum of Natural History Brussels	Libya	30,766	17,833	1	Short			
16995	Belgium Royal Museum of Natural	Libya	21,963	24,820	2	Short			

	History Brussels								
16994	Belgium Royal Museum of Natural History Brussels	Libya	21,963	24,820	2	Short			
11908	Natural History Museum (Naturhistorisches Museum) – Vienna	Tunisia	34,466	8,716					
ZBSC 0169	Morocco 2011	Morocco	32,894316	-4,999255	outgroup	Short			
ZBSC 0170	Morocco 2011	Morocco	32.894316	-4.999255	outgroup	Short/Long	*ADRA2B, GHR, IRBP, ΰWF, DBX5*		
ZBSC 0171	Morocco 2011	Morocco	32.895635	-5.003515	outgroup	Short/Long			
ZBSC 0172	Morocco 2011	Morocco	32.895635	-5.003515	outgroup	Short			
ZBSC 0173	Morocco 2011	Morocco	32.895635	-5.003515	outgroup	Short			
ZBSC 0174	Morocco 2011	Morocco	32.910272	-5.032408	outgroup	Short/Long	*ADRA2B, GHR, IRBP, ΰWF, DBX5*		
ZBSC 0400	Morocco 2011	Morocco	33.25945	2.69051	outgroup	Short/Long	*ADRA2B, GHR, IRBP, ΰWF, DBX6*		
10178	Morocco 2012	Morocco	34,388593	-3,009135	outgroup	Short/Long	*ADRA2B, GHR, IRBP, ΰWF, DBX7*		
10233	Morocco 2013	Morocco	32,939682	-2,490725	outgroup	Short/Long	*ADRA2B, GHR, IRBP, ΰWF, DBX8*		
10234	Morocco 2013	Morocco	32,820313	-2,585565	outgroup	Short/Long	*ADRA2B, GHR, IRBP, ΰWF, DBX9*		

ADRA2B, *alpha-2B adrenergic receptor;* ***GHR****, growth hormone receptor;* ***IRBP***, *interstitial retinoid binding protein*); ***ΰWF*** (*von Willebrand factor*); ***DBX5***, intron 5 from the *DBX* gene; ***M,*** microsatellites.

[a] Boratyński *et al.*2012
[b] Boratyński *et al.*2014

Annex 2. Details of loci, PCR primers (primers sequences in 5'-3') and respective conditions used.

Locus	Size(bp)	Forward Primer		Reverse primer		Ta (°C)	Reference
Cytb	897	Jac1Fw	GGACTCCCCATGACCTAT	Jac1Rv	TGCTGGTTTACAAGACCA	55*	Boratyński *et al.*2012
		Jac4Fw	CAAACCCACTTAATACGC	Jac4Rv	CGAGAAGAGGGATACGAC		
	325	Jac4Fw	CAAACCCACTTAATACGC	Jac1Rv	TGCTGGTTTACAAGACCA		
DBX5	317	DBX5F	CAACAACTGTCCTCCACA	DBX5R	CATGATAATTTCTCCCATCTC	TD 60-50 (-0.5)	Hellborg & Ellegren 2004
ADRA2B	693	J-adra2B1F	CTGGCGCTCGACGTGCTCTT	J-adra2B1R	AGCACCTGGCCACGGAGAGT	TD 68-64 (-0.5)	This study
GHR	798	J-ghr3F	ACAATGATGACTCTTGGGTTGAGT	J-ghr3R	AAGGGCAGGGCAGTTGCATT		This study
IRBP	1058	J-irbp2F	GCGGCCATCCAGCAGGTAAT	J-irbp3R	CCGGCAGCACTGACACCTGA		This study
ΰWF	874	J-ΰWF4F	ACGGATGCCTCGCTCAGCTC	J-ΰWF7R	CTCCAGTTCCTGCTGGTTGGCA		Boratyński *et al.*2012

*Differences in the duration of the extension step (1.15min-long fragment; 45s-short fragment); Ta (°C) is the annealing temperature used in the PCR (where "TD" represents a touchdown procedure).

Annex 3. List of the GenBank sequences used in the study.

Identification	GB Accession number	Province, Country	Lat	Long	Clade	Cytb Sequence	Reference
D144	JX885131	Richard Toll, Senegal	16,43	-15,66	1	Short/Long	Faleh *et. al* 2012
D145	JX885132	Thilé, Senegal	16,51	-15,08	1	Short/Long	Faleh *et. al* 2012
D146	JX885133	Nbeika, Mauritania	17,95	-12,23	1	Short/Long	Faleh *et. al* 2012
D147	JX885134	Nbeika, Mauritania	17,95	-12,23	1	Short/Long	Faleh *et. al* 2012
D149	JX885136	Akjoujt, Mautitania	19,74	-14,37	1	Short/Long	Faleh *et. al* 2012
D150	JX885137	Ayn El taya, Mauritania	20,25	-13,27	1	Short/Long	Faleh *et. al* 2012
D151	JX885138	Ayn El taya, Mauritania	20,25	-13,27	2	Short/Long	Faleh *et. al* 2012
D152	JX885139	El Mhaoudat, Mauritania	22,97	-12,00	2	Short/Long	Faleh *et. al* 2012
D153	JX885140	El Mhaoudat, Mauritania	22,97	-12,00	2	Short/Long	Faleh *et. al* 2012
D154	JX885141	Niafounke, Mali	15,93	-3,97	1	Short/Long	Faleh *et. al* 2012
D155	JX885142	Niafounke , Mali	15,93	-3,97	1	Short/Long	Faleh *et. al* 2012
D156	JX885143	Tsinsack, Mali	16,73	-2,96	1	Short/Long	Faleh *et. al* 2012
D157	JX885144	Tsinsack, Mali	16,73	-2,96	1	Short/Long	Faleh *et. al* 2012
D158	JX885145	Tidarmene, Mali	17,02	2,11	1	Short/Long	Faleh *et. al* 2012
D159	JX885146	Tidarmene, Mali	17,02	2,11	1	Short/Long	Faleh *et. al* 2012
D160	JX885147	Abeibara, Mali	19,01	1,83	2	Short/Long	Faleh *et. al* 2012
D161	JX885148	Abeibara, Mali	19,01	1,83	2	Short/Long	Faleh *et. al* 2012
D162	JX885149	Oued Chacheguerène, Mali	19,70	0,00	1	Short/Long	Faleh *et. al* 2012
D163	JX885150	Oued Chacheguerène, Mali	19,70	0,00	2	Short/Long	Faleh *et. al* 2012
D164	JX885151	Kreb in Karoua, Mali	19,78	0,33	1	Short/Long	Faleh *et. al* 2012
D165	JX885152	Babangata, Niger	12,91	2,40	1	Short/Long	Faleh *et. al* 2012
D166	JX885153	Gangara, Niger	14,61	8,50	1	Short/Long	Faleh *et. al* 2012
D168	JX885155	Gangara, Niger	14,61	8,50	2	Short/Long	Faleh *et. al* 2012
D169	JX885156	Gangara, Niger	14,61	8,50	2	Short/Long	Faleh *et. al* 2012
D170	JX885157	Gangara , Niger	14,61	8,50	1	Short/Long	Faleh *et. al* 2012
D171	JX885158	Goûgaram, Niger	18,55	7,78	2	Short/Long	Faleh *et. al* 2012
D172	JX885159	Al Baydah, Algeria	33,69	1,01	1	Short/Long	Faleh *et. al* 2012
D174	JX885161	Batna, Algeria	35,32	5,83	2	Short/Long	Faleh *et. al* 2012
D175	JX885162	Tebessa, Algeria	35,40	8,12	2	Short/Long	Faleh *et. al* 2012
D176	JX885163	Tebessa, Algeria	35,40	8,12	2	Short/Long	Faleh *et. al* 2012
D191	JX885164	Hamma, Tunisia	33,95	9,63	2	Short/Long	Faleh *et. al* 2012
D194	JX885165	Matmata, Tunisia	33,55	9,96	2	Short/Long	Faleh *et. al* 2012
D177	JX885166	Sbeitla, Tunisia	35,22	9,12	1	Short/Long	Faleh *et. al* 2012
D178	JX885167	Sbeitla, Tunisia	35,22	9,12	1	Short/Long	Faleh *et. al* 2012
D179	GU433437	Sbeitla, Tunisia	35,22	9,12	2	Short/Long	Faleh *et. al* 2010
D180	GU433439	Menzel Chaker, Tunisia	34,96	10,36	2	Short	Faleh *et. al* 2010
D182	GU433422	Nefta, Tunisia	33,87	7,87	1	Short	Faleh *et. al* 2010
D186	GU433408	Nefta, Tunisia	33,87	7,87	1	Short	Faleh *et. al* 2010

D187	GU433409	Douz, Tunisia	33,45	9,02	1	Short	Faleh *et. al* 2010
D188	GU433410	Douz, Tunisia	33,45	9,02	1	Short	Faleh *et. al* 2010
D189	GU433424	Douz, Tunisia	33,45	9,02	1	Short/Long	Faleh *et. al* 2010
D190	GU433412	Hamma, Tunisia	33,95	9,63	1	Short	Faleh *et. al* 2010
D192	GU433413	Hamma, Tunisia	33,95	9,63	1	Short	Faleh *et. al* 2010
D195	GU433415	Matmata, Tunisia	33,55	9,96	1	Short	Faleh *et. al* 2010
D197	GU433427	Matmata, Tunisia	33,55	9,96	1	Short/Long	Faleh *et. al* 2010
D198	GU433411	Matmata, Tunisia	33,55	9,96	1	Short	Faleh *et. al* 2010
D200	GU433421	Matmata, Tunisia	33,55	9,96	1	Short	Faleh *et. al* 2010
D201	GU433441	Matmata, Tunisia	33,55	9,96	2	Short/Long	Faleh *et. al* 2010
D202	GU433418	Matmata, Tunisia	33,55	9,96	1	Short/Long	Faleh *et. al* 2010
D203	GU433435	Matmata, Tunisia	33,55	9,96	2	Short/Long	Faleh *et. al* 2010
D205	GU433438	Matmata, Tunisia	33,55	9,96	2	Short	Faleh *et. al* 2010
D207	GU433425	Matmata, Tunisia	33,55	9,96	1	Short/Long	Faleh *et. al* 2010
D208	GU433426	Matmata, Tunisia	33,55	9,96	1	Short/Long	Faleh *et. al* 2010
D212	GU433429	Matmata, Tunisia	33,55	9,96	1	Short/Long	Faleh *et. al* 2010
D215	GU433420	Tataouine, Tunisia	32,93	10,45	1	Short/Long	Faleh *et. al* 2010
D209	GU433423	Matmata, Tunisia	33,55	9,96	1	Short/Long	Faleh *et. al* 2010
D213	JX885168	Medenine, Tunisia	33,33	11,00	2	Short	Faleh *et. al* 2012
D217	JX885169	Hun, Libya	29,12	15,93	2	Short/Long	Faleh *et. al* 2012
D218	JX885170	Hun, Libya	29,12	15,93	2	Short/Long	Faleh *et. al* 2012
D219	JX885171	Hun, Libya	29,12	15,93	1	Short/Long	Faleh *et. al* 2012
D220	JX885172	Hun, Libya	29,12	15,93	1	Short/Long	Faleh *et. al* 2012
D221	JX885173	Adiri, Libya	27,53	13,19	1	Short/Long	Faleh *et. al* 2012
D222	JX885174	Adiri, Libya	27,53	13,19	1	Short/Long	Faleh *et. al* 2012
D223	JX885175	Adiri, Libya	27,53	13,19	1	Short/Long	Faleh *et. al* 2012
D224	JX885176	Adiri, Libya	27,53	13,19	1	Short	Faleh *et. al* 2012
D225	JX885177	Birak, Libya	27,54	14,24	1	Short/Long	Faleh *et. al* 2012
D226	JX885178	Birak, Libya	27,54	14,24	1	Short/Long	Faleh *et. al* 2012
D227	JX885179	Birak, Libya	27,54	14,24	1	Short/Long	Faleh *et. al* 2012
D228	JX885180	Birak, Libya	27,54	14,24	1	Short/Long	Faleh *et. al* 2012
D229	JX885181	Jalu, Libya	29,02	21,55	1	Short/Long	Faleh *et. al* 2012
D230	JX885182	Jalu, Libya	29,02	21,55	1	Short/Long	Faleh *et. al* 2012
D231	JX885183	Tazirbu, Libya	25,70	21,13	1	Short/Long	Faleh *et. al* 2012
D232	JX885184	Tazirbu, Libya	25,70	21,13	2	Short/Long	Faleh *et. al* 2012
D233	JX885185	Burj El Arab, Egypt	29,97	31,27	2	Short/Long	Faleh *et. al* 2012
D234	JX885186	Burj El Arab, Egypt	29,97	31,27	2	Short/Long	Faleh *et. al* 2012
D235	JX885187	Burj El Arab, Egypt	29,97	31,27	2	Short/Long	Faleh *et. al* 2012
D236	JX885188	Abu Rauwash , Egypt	29,66	31,23	2	Short/Long	Faleh *et. al* 2012
D237	JX885189	Al Jizah, Egypt	29,15	29,98	2	Short/Long	Faleh *et. al* 2012
D239	JX885191	Al Minya, Egypt	28,17	30,74	2	Short/Long	Faleh *et. al* 2012
D240	JX885192	Al Minya, Egypt	28,17	30,74	2	Short/Long	Faleh *et. al* 2012
D241	JX885193	Sulaibiya, Kuwait	29,27	47,71	2	Short/Long	Faleh *et. al* 2012
D242	JX885194	Sulaibiya, Kuwait	29,27	47,71	2	Short/Long	Faleh *et. al* 2012
USNM483105	KC663576	Morocco	28,77	-10,23	2	Short	Boratyński *et al.* 2014
USNM482686	KC663514	Niger	15,75	6,60	2	Short	Boratyński *et al.* 2014

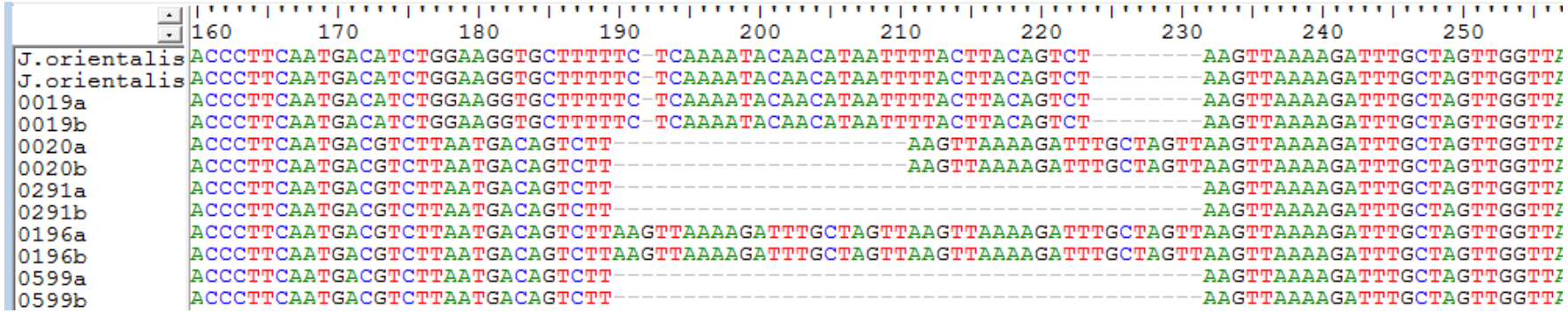
USNM482681	KC663512	Niger	17,37	6,72	1	Short	Boratyński *et al.* 2014
USNM482673	KC663513	Niger	18,97	5,97	1	Short	Boratyński *et al.* 2014
USNM482671	KC663526	Niger	16,55	6,87	2	Short	Boratyński *et al.* 2014
USNM482504	KC663580	Algeria	23,17	5,12	1	Short	Boratyński *et al.* 2014
USNM482503	KC663577	Algeria	22,93	5,42	2	Short	Boratyński *et al.* 2014
USNM482502	KC663565	Algeria	22,63	5,73	2	Short	Boratyński *et al.* 2014
USNM482499	KC663557	Algeria	23,57	5,12	1	Short	Boratyński *et al.* 2014
USNM482491	KC663549	Algeria	26,87	-0,97	1	Short	Boratyński *et al.* 2014
USNM482482	KC663539	Algeria	30,05	-2,22	1	Short	Boratyński *et al.* 2014
USNM482480	KC663531	Algeria	32,46	-0,58	1	Short	Boratyński *et al.* 2014
USNM475885	KC663572	Morocco	32,68	-3,08	2	Short	Boratyński *et al.* 2014
USNM475865	KC663579	Morocco	30,30	-5,93	2	Short	Boratyński *et al.* 2014
USNM475820	KC663525	Morocco	32,15	-1,25	1	Short	Boratyński *et al.* 2014
USNM475797	KC663524	Morocco	31,83	-4,58	1	Short	Boratyński *et al.* 2014
USNM475783	KC663574	Morocco	32,50	-2,05	2	Short	Boratyński *et al.* 2014
USNM475780	KC663575	Morocco	31,95	-3,55	1	Short	Boratyński *et al.* 2014
USNM475764	KC663573	Morocco	32,12	-2,85	2	Short	Boratyński *et al.* 2014
USNM475761	KC663578	Morocco	31,90	-4,48	2	Short	Boratyński *et al.* 2014
USNM401212	KC663571	Mauritania	21,52	-13,05	1	Short	Boratyński *et al.* 2014
USNM350066	KC663536	Egypt	30,08	31,58	2	Short	Boratyński *et al.* 2014
USNM342084	KC663515	Sudan	15,23	36,39	1	Short	Boratyński *et al.* 2014
USNM342040	KC663521	Egypt	29,70	32,35	1	Short	Boratyński *et al.* 2014
USNM342034	KC663529	Egypt	28,54	30,57	1	Short	Boratyński *et al.* 2014
USNM342033	KC663540	Egypt	28,32	31,12	2	Short	Boratyński *et al.* 2014
USNM342030	KC663528	Egypt	27,14	31,38	2	Short	Boratyński *et al.* 2014
USNM342028	KC663543	Egypt	27,22	30,80	1	Short	Boratyński *et al.* 2014
USNM325828	KC663550	Libya	29,59	24,86	1	Short	Boratyński *et al.* 2014
USNM325821	KC663569	Libya	31,19	16,40	1	Short	Boratyński *et al.* 2014
USNM325819	KC663566	Libya	30,55	18,47	2	Short	Boratyński *et al.* 2014
USNM325805	KC663563	Libya	29,57	24,70	1	Short	Boratyński *et al.* 2014
USNM325802	KC663562	Libya	29,75	24,55	1	Short	Boratyński *et al.*

							2014
USNM325789	KC663570	Libya	25,75	21,15	1	Short	Boratyński *et al.* 2014
USNM325774	KC663561	Libya	29,25	21,23	1	Short	Boratyński *et al.* 2014
USNM325770	KC663560	Libya	32,42	13,05	1	Short	Boratyński *et al.* 2014
USNM322811	KC663554	Libya	27,53	13,20	1	Short	Boratyński *et al.* 2014
USNM322809	KC663553	Libya	27,55	14,25	1	Short	Boratyński *et al.* 2014
USNM322803	KC663568	Libya	27,00	14,45	1	Short	Boratyński *et al.* 2014
USNM322798	KC663558	Libya	25,90	13,89	1	Short	Boratyński *et al.* 2014
USNM322788	KC663556	Libya	24,95	10,21	1	Short	Boratyński *et al.* 2014
USNM322770	KC663555	Libya	26,77	14,00	1	Short	Boratyński *et al.* 2014
USNM322767	KC663559	Libya	27,22	14,66	1	Short	Boratyński *et al.* 2014
USNM322762	KC663564	Libya	29,09	15,90	1	Short	Boratyński *et al.* 2014
USNM321864	KC663551	Libya	30,75	11,52	2	Short	Boratyński *et al.* 2014
USNM321863	KC663567	Libya	32,06	11,35	1	Short	Boratyński *et al.* 2014
USNM319773	KC663552	Libya	24,18	23,32	1	Short	Boratyński *et al.* 2014
USNM317068	KC663546	Egypt	28,56	33,96	2	Short	Boratyński *et al.* 2014
USNM317065	KC663527	Egypt	24,01	32,83	1	Short	Boratyński *et al.* 2014
USNM317059	KC663522	Egypt	30,41	30,60	2	Short	Boratyński *et al.* 2014
USNM317050	KC663534	Egypt	30,10	31,58	1	Short	Boratyński *et al.* 2014
USNM317049	KC663519	Egypt	30,63	29,84	1	Short	Boratyński *et al.* 2014
USNM317047	KC663518	Egypt	29,49	30,40	1	Short	Boratyński *et al.* 2014
USNM317041	KC663520	Egypt	30,22	30,90	1	Short	Boratyński *et al.* 2014
USNM317028	KC663541	Egypt	25,26	32,46	1	Short	Boratyński *et al.* 2014
USNM317020	KC663537	Egypt	30,50	30,79	2	Short	Boratyński *et al.* 2014
USNM317018	KC663533	Egypt	30,41	30,60	1	Short	Boratyński *et al.* 2014
USNM317017	KC663542	Egypt	25,67	32,77	1	Short	Boratyński *et al.* 2014
USNM317015	KC663544	Egypt	31,52	25,61	2	Short	Boratyński *et al.* 2014
USNM317014	KC663532	Egypt	22,30	36,54	1	Short	Boratyński *et al.* 2014
USNM317013	KC663545	Egypt	22,27	36,40	1	Short	Boratyński *et al.* 2014
USNM317012	KC663530	Egypt	22,53	36,23	1	Short	Boratyński *et al.* 2014
USNM297613	KC663516	Sudan	19,87	37,18	1	Short	Boratyński *et al.* 2014

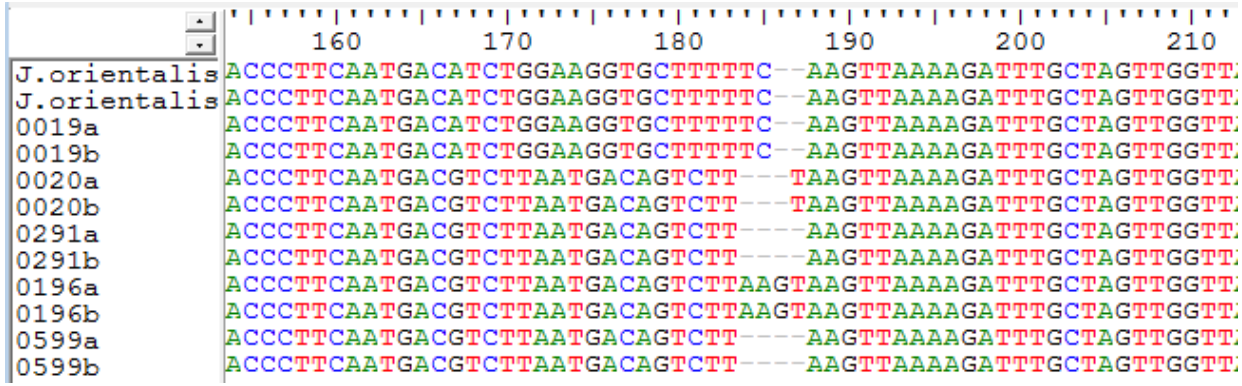
USNM297612	KC663517	Sudan	19,54	37,19	1	Short	Boratyński *et al.* 2014
USNM283260	KC663538	Egypt	30,31	32,28	2	Short	Boratyński *et al.* 2014
USNM282539	KC663535	Egypt	30,09	31,43	2	Short	Boratyński *et al.* 2014
2002512	JX885154	Babangata, Niger	12,91	2,40	2	Short/Long	Faleh *et. al* 2012
2002274	JX885160	Goûgaram, Niger	18,55	7,78	1	Short/Long	Faleh *et. al* 2012

Annex 4. Sequence alignment of the DBX intron wherein the insertion/deletion polymorphisms can be observed (a). Indels were coded so that a more robust approximation of the mutational steps could be made for the phylogenetic networks reconstruction. In (b) is displayed the alignment after coding the polymorphisms.

(a)



(b)



Annex 5. Evolutionary models of each locus used for phylogenetic and demographic analyses. Calculations were made with jModelTest (Darriba *et al.* 2012).

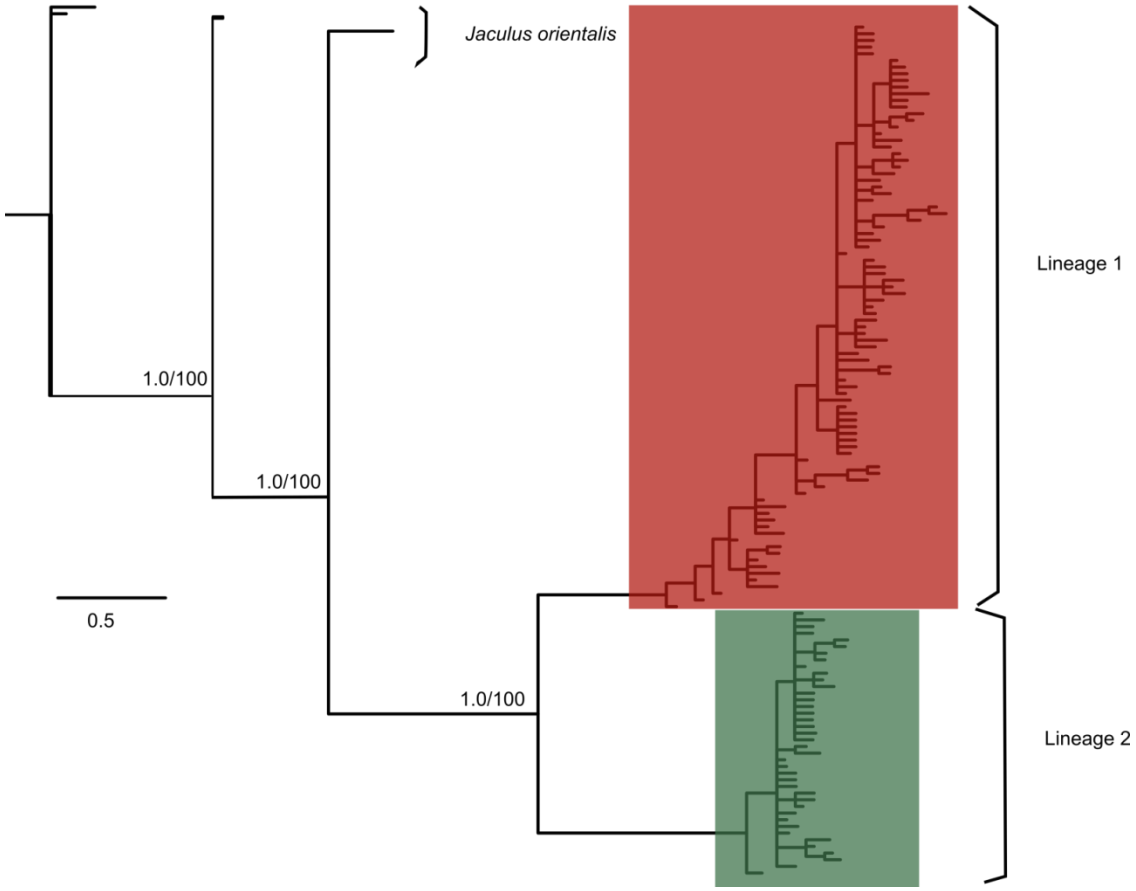
	Phylogenetics analyses/ Species tree inference	**EBSP analyses**	
		Lineage 1	**Lineage 2**
Cytb	HKY+I+G	HKY+I+G	HKY+I
DBX5	HKY	-	-
***DBX5* without recombination**	HKY	HKY	HKY
ADRA2B	HKY+I	HKY+I	HKY+I
GHR	HKY	HKY	HKY
IRBP	HKY+I	-	-
***IRBP* without recombination**	HKY	HKY+I	HKY+I
ʊWF	GTR+I+G	-	-
***ʊWF* without recombination**	HKY+I	HKY+I	HKY+I

Annex 6. List of the multiplexes and the primer characteristics (primers sequences in 5'-3') used in microsatellite analysis.

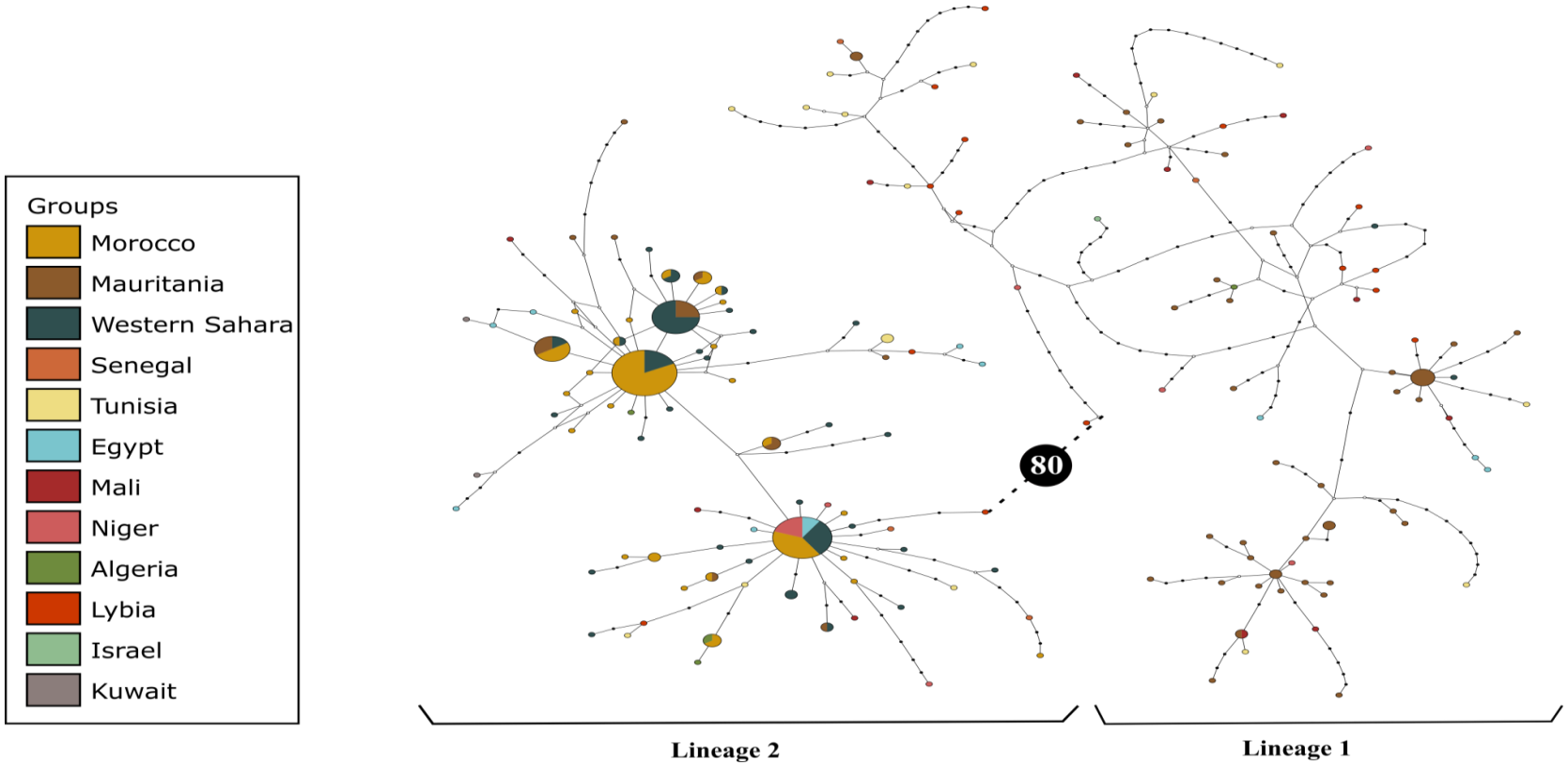
Multiplex	Locus	Size(bp)	Forward Primer	Reverse primer	Repeat number	Motif	Volume (in 100µL)	Tail	Ta (°C)	Nº of cycles
1	**Jac01**	101	GATGGCTGTAGCTGTCTGGG	GAACCATAGTAAGATAACAGCATGG	14	tg	0.8	FAM	TD 57-51 (-0.5)	40
	Jac02	144	CACAGACTGAAACCGTGAGC	CCAAAGAGGAGGCACAGAAG	12	ac	0.8	FAM		
	Jac04	105	ATCAGCCTCTCAGCCTTCTG	ACTGCAGGCTCTCGTGTTCT	11	ga	0.8	VIC		
	Jac23	234	AACAAGAATGAATACATGGGGA	TAGGTGTGCACCACCACACT	11	ac	1.4	VIC		
	Jac07	95	TTCATGCCAAGTTCAAAGGC	ATCGCAACAAGAAAGATGGC	18	ac	0.8	NED		
	Jac08	140	CAAGGAACGTGCCTGACTTT	TAGCGTCCCTGTTTTCCTTC	12	ac	2.0	NED		
	Jac24	141	ACAGTCCCCTTTAACATGATAGTC	CTTCTGTTAGTAGCTGAGACATGATT	16	gt	2.0	PET		
2	**Jac11**	111	CCACCTTCTATCATAAATACACAGTGA	GGCCGTTGTATGTGAGTCAA	21	ca	2.0	FAM	TD 55-49 (-0.5)	45
	Jac27	140	GGTGTAACCCTGACCTAATCC	TGTCTATGTAACTCATGACCAAGAA	14	ac	2.0	VIC		
	Jac16	190	TCTGTCTTAGGAATATTGGGCA	TGTCTTGATTTCTTCTCTGTTTTATTG	12	ac	0.8	VIC		
	Jac18	176	GCCCCAATATTTCATGTTTCA	GGCTTCTGGAGTTCATTTGC	20	ca	1.0	NED		
Individual reactions	**Jac12**	166	ACCTGCCAGCAACGATGT	GGCCGTTGTATGTGAGTCAA	20	tg	-	FAM	TD 58-51 (-0.5)	40
	Jac37	190	TGTCACATGAAATTAATAGGGCAT	TCTTTGGTATTCCTCAACTCGG	11	atgaa	-	PET		

Ta (°C) is the annealing temperature used in the PCR (where "TD" represents a touchdown procedure); the Nº of cycles corresponds to the total number of amplification cycles in the PCR reaction; all primers were produced in this study.

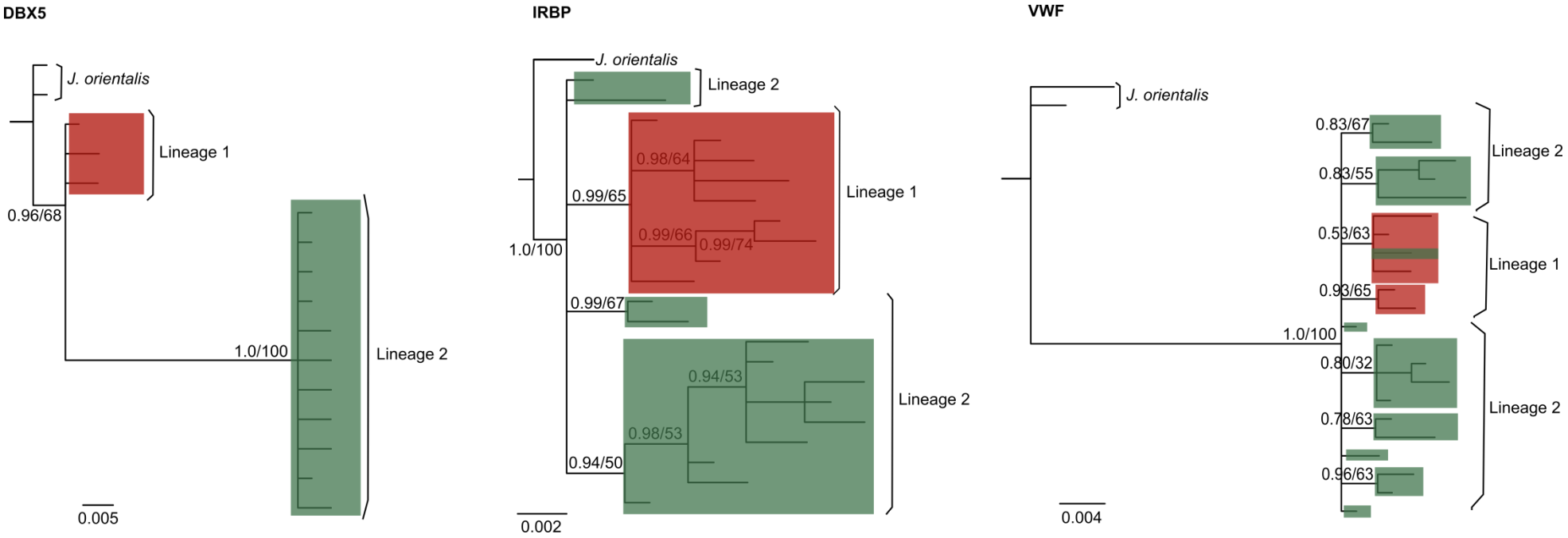
Annex 7. Phylogenetic tree based on Bayesian inference showing the relationship among the two lineages in *J. jaculus* for the short fragment of cytochrome b (*cytb*) gene (n=325). Values on branches indicate posterior probability support and bootstrap values of the maximum-likelihood analysis, respectively. On each clade is indicated its respective lineage (red – Lineage 1; green – Lineage 2). *J. orientalis* (n=10) was used as outgroup.



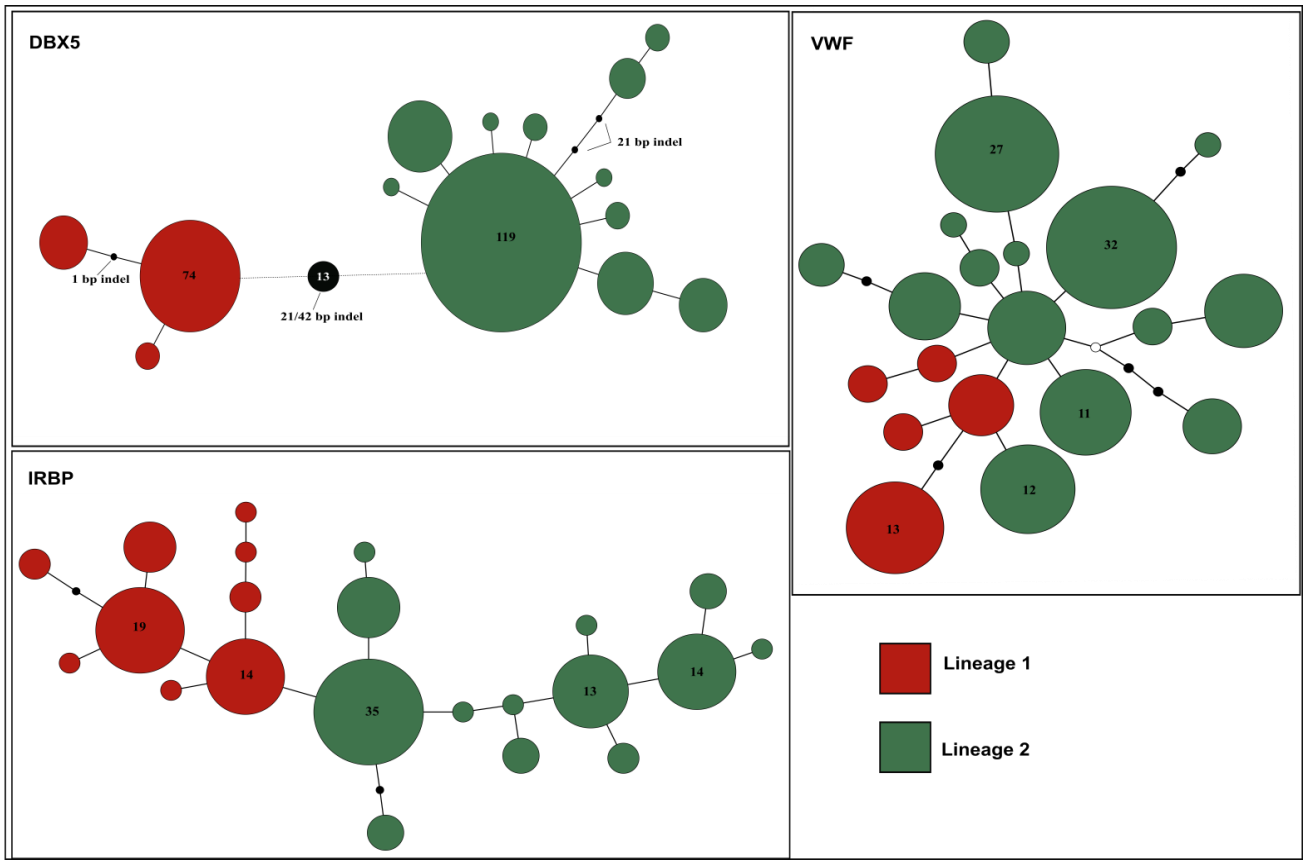
Annex 8. Statistical parsimony haplotype network of the long fragment of cytochrome *b* gene of *Jaculus jaculus* specimens (n=210). Each circle represents one haplotype and the circle area is proportional to the frequency of each haplotype. The size of the branches is proportional to the number of nucleotide differences between haplotypes, and dots on branches specify mutational steps. The number of mutational steps between the two lineages is specified on the figure (80). Haplotypes are coloured according to the sampling region (see legend). The network was edited in tcsBU (Santos *et al.* 2015).



Annex 9. Phylogenetic trees based on Bayesian inference showing the relationships among *J. jaculus* specimens for the X-chromosome intron (*DBX5*) and nuclear autosomal genes (*IRBP* and *υWF*) without the recombinants blocks of the dataset. Number of sequences used for each locus is specified in Table 1. Values on branches designate the posterior probability support and bootstrap values for the maximum-likelihood approach, respectively. The two colours are in agreement with the two mitochondrial lineages (red – Lineage 1; green – Lineage 2). *J. orientalis* (n=6) was used as outgroup.



Annex 10. Statistical parsimony haplotype networks of the X-Chromosome intron (*DBX5*) and nuclear autosomal genes (*IRBP* and *υWF*) without the recombinants blocks of the dataset (number of sequences used for each locus is specified in Table 1). Each circle represents one haplotype and the circle area is proportional to the frequency of each haplotype. Total frequencies are indicated for more common haplotypes. The size of the branches is proportional to the number of nucleotide differences between haplotypes, and dots on branches specify mutational steps. The insertion/deletion polymorphisms (indels) of the *DBX* intron were coded as single mutational steps (see Annex 4) and so the sizes of the indels are indicated on the respective mutational step. The number of mutational steps between the two lineages is specified for the *DBX* gene (13). Haplotypes are coloured according to the respective mitochondrial lineage.



Annex 11. Summary of the calculations performed for the IM estimations of effective population sizes, divergence time and population migration rate. The geometric means calculated from Dxy and Da between *J. orientalis* (J.o) and *J. jaculus* (J.j) is presented.

	Dxy(J.o/J.j)	Dxy/2	Div(J.o/J.a)Myr	DivMyr*ngen	μ	L	u
Cytb	0,131	0,0655	5,97	5970000	1,09715E-08	892	1,1986E-06
DBX5	0,025	0,0125	5,97	5970000	2,0938E-09	693	9,31199E-07
ADRA2B	0,008	0,004	5,97	5970000	6,70017E-10	798	1,07229E-06
GHR	0,007	0,0035	5,97	5970000	5,86265E-10	681	9,15075E-07
IRBP	0,003	0,0015	5,97	5970000	2,51256E-10	514	6,90673E-07
υWF	0,031	0,0155	5,97	5970000	2,59631E-09	295	3,96398E-07
Geometric mean					1,34372E-09		8,18116E-07

	Da(J.o/J.a)	Da/2	Div(J.o/J.a)Mya	DivMya*ngen	μ	L	u
Cytb	0,099	0,0495	5,97	5970000	8,29146E-09	892	7,15197E-07
DBX5	0,018	0,009	5,97	5970000	1,50754E-09	693	5,55641E-07
ADRA2B	0,004	0,002	5,97	5970000	3,35008E-10	798	6,39829E-07
GHR	0,004	0,002	5,97	5970000	3,35008E-10	681	5,46019E-07
IRBP	0,001	0,0005	5,97	5970000	8,37521E-11	514	4,1212E-07
υWF	0,027	0,0135	5,97	5970000	2,26131E-09	295	2,36528E-07
Geometric mean					8,01791E-10		4,88165E-07

Dxy - Average number of nucleotide substitutions per site between J.o and J.j

Da - Number of net nucleotide substitutions per site between J.o and J.j

Div(J.o/J.j) – Divergence in Myr between J.o and J.j (5.97 Myr; Pisano *et al.* 2015)

Ngen – number of generations (ngen=1) [Nabholz *et al.* 2008]

μ - substitutions/site/generation (estimated mutation rate)

L – Length of the fragment of each locus

u - substitutions/fragment/generation (geometric mean used for further calculations)

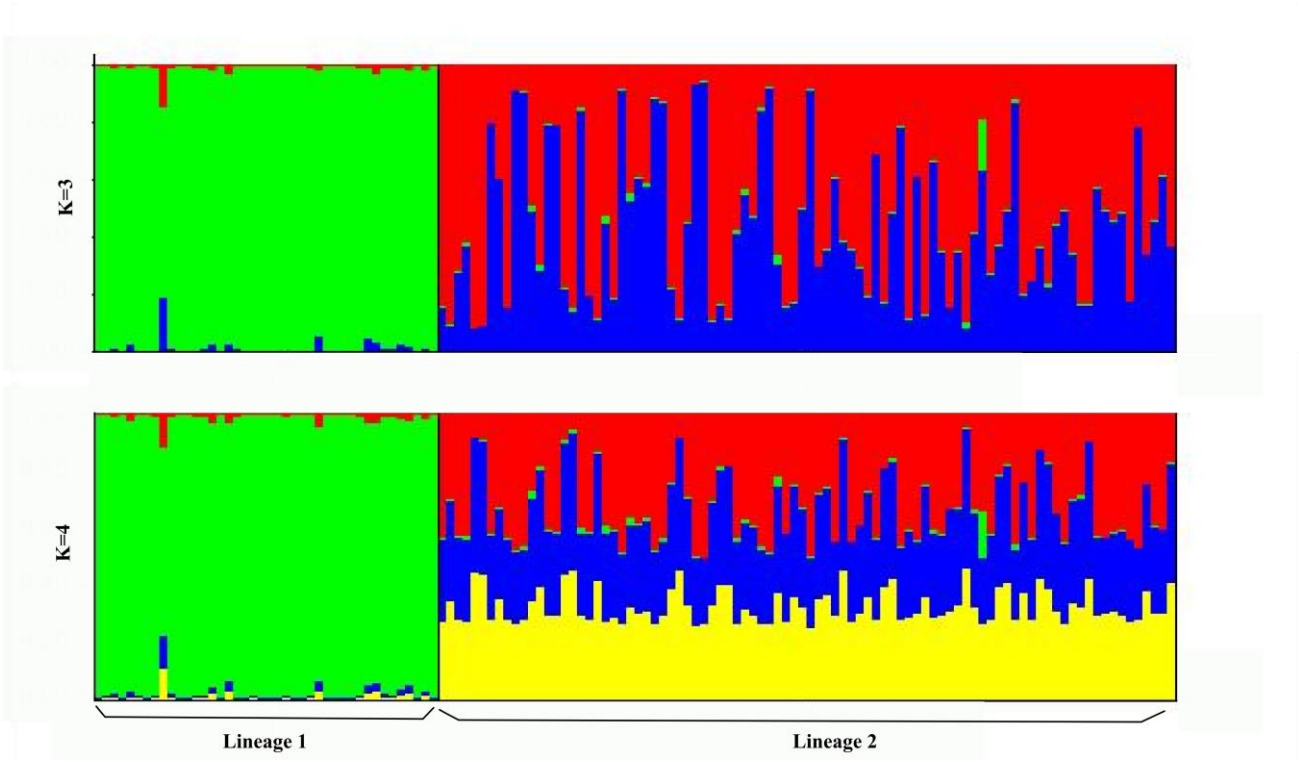
Annex 12. List of the genotyped samples, assigned cluster and proportion of membership estimated by Structure. Highlighted are the two samples that revealed deviations from the expected proportion of membership to the respective cluster.

Sample code	Cluster	Proportion of membership	
		L1	L2
11043	Lineage 1	0.995	0.005
11379	Lineage 1	0.994	0.006
11382	Lineage 1	0.994	0.006
11384	Lineage 1	0.996	0.004
11408	Lineage 1	0.978	0.022
6064	Lineage 1	0.995	0.005
6268	Lineage 1	0.995	0.005
6269	Lineage 1	0.991	0.009
6367	Lineage 1	0.795	0.205
D1003	Lineage 1	0.992	0.008
D101	Lineage 1	0.995	0.005
D113	Lineage 1	0.994	0.006
D22	Lineage 1	0.991	0.009
D3055	Lineage 1	0.993	0.007
D3107	Lineage 1	0.984	0.016
D320	Lineage 1	0.995	0.005
D493	Lineage 1	0.949	0.051
D511	Lineage 1	0.995	0.005
D52	Lineage 1	0.994	0.006
D53	Lineage 1	0.995	0.005
D536	Lineage 1	0.993	0.007
D541	Lineage 1	0.995	0.005
ZBSC0019	Lineage 1	0.996	0.004
ZBSC0021	Lineage 1	0.994	0.006
ZBSC0064	Lineage 1	0.996	0.004
ZBSC0072	Lineage 1	0.997	0.003
ZBSC0079	Lineage 1	0.994	0.006
ZBSC0218	Lineage 1	0.972	0.028
ZBSC0241	Lineage 1	0.995	0.005
ZBSC0243	Lineage 1	0.995	0.005
ZBSC0244	Lineage 1	0.995	0.005
ZBSC0256	Lineage 1	0.996	0.004
ZBSC0265	Lineage 1	0.991	0.009
ZBSC0303	Lineage 1	0.980	0.020
ZBSC0417	Lineage 1	0.978	0.022
ZBSC0420	Lineage 1	0.995	0.005
ZBSC0509	Lineage 1	0.989	0.011
ZBSC0526	Lineage 1	0.972	0.028
ZBSC0542	Lineage 1	0.975	0.025
ZBSC0543	Lineage 1	0.995	0.005

ZBSC0558	Lineage 1	0.992	0.008
ZBSC0559	Lineage 1	0.996	0.004
10355	Lineage 2	0.004	0.996
11411	Lineage 2	0.005	0.995
6481	Lineage 2	0.005	0.995
6482	Lineage 2	0.005	0.995
8067	Lineage 2	0.005	0.995
8069	Lineage 2	0.004	0.996
8070	Lineage 2	0.005	0.995
8072	Lineage 2	0.008	0.992
8076	Lineage 2	0.004	0.996
8345	Lineage 2	0.004	0.996
8350	Lineage 2	0.012	0.988
8367	Lineage 2	0.021	0.979
8374	Lineage 2	0.021	0.979
9072	Lineage 2	0.007	0.993
9083	Lineage 2	0.004	0.996
9086	Lineage 2	0.011	0.989
9087	Lineage 2	0.019	0.981
D100	Lineage 2	0.005	0.995
D1283	Lineage 2	0.004	0.996
D144	Lineage 2	0.005	0.995
D145	Lineage 2	0.006	0.994
D549	Lineage 2	0.005	0.995
D576	Lineage 2	0.006	0.994
D577	Lineage 2	0.034	0.966
D578	Lineage 2	0.009	0.991
D800	Lineage 2	0.016	0.984
ZBSC0020	Lineage 2	0.005	0.995
ZBSC0070	Lineage 2	0.007	0.993
ZBSC0081	Lineage 2	0.009	0.991
ZBSC0082	Lineage 2	0.006	0.994
ZBSC0083	Lineage 2	0.013	0.987
ZBSC0084	Lineage 2	0.006	0.994
ZBSC0193	Lineage 2	0.009	0.991
ZBSC0196	Lineage 2	0.005	0.995
ZBSC0197	Lineage 2	0.010	0.990
ZBSC0198	Lineage 2	0.005	0.995
ZBSC0219	Lineage 2	0.011	0.989
ZBSC0224	Lineage 2	0.021	0.979
ZBSC0226	Lineage 2	0.004	0.996
ZBSC0240	Lineage 2	0.020	0.980
ZBSC0242	Lineage 2	0.019	0.981
ZBSC0245	Lineage 2	0.035	0.965
ZBSC0291	Lineage 2	0.004	0.996
ZBSC0292	Lineage 2	0.004	0.996

ZBSC0293	Lineage 2	0.007	0.993
ZBSC0294	Lineage 2	0.010	0.990
ZBSC0295	Lineage 2	0.007	0.993
ZBSC0296	Lineage 2	0.005	0.995
ZBSC0306	Lineage 2	0.004	0.996
ZBSC0383	Lineage 2	0.007	0.993
ZBSC0384	Lineage 2	0.005	0.995
ZBSC0385	Lineage 2	0.005	0.995
ZBSC0388	Lineage 2	0.005	0.995
ZBSC0403	Lineage 2	0.004	0.996
ZBSC0404	Lineage 2	0.005	0.995
ZBSC0408	Lineage 2	0.008	0.992
ZBSC0413	Lineage 2	0.006	0.994
ZBSC0414	Lineage 2	0.004	0.996
ZBSC0423	Lineage 2	0.008	0.992
ZBSC0424	Lineage 2	0.004	0.996
ZBSC0591	Lineage 2	0.009	0.991
ZBSC0599	Lineage 2	0.005	0.995
ZBSC0614	Lineage 2	0.005	0.995
ZBSC0626	Lineage 2	0.009	0.991
ZBSC0629	Lineage 2	0.023	0.977
ZBSC0647	Lineage 2	0.006	0.994
ZBSC0650	Lineage 2	0.246	0.754
ZBSC0651	Lineage 2	0.005	0.995
ZBSC0652	Lineage 2	0.007	0.993
ZBSC0653	Lineage 2	0.006	0.994
ZBSC0654	Lineage 2	0.024	0.976
ZBSC0665	Lineage 2	0.005	0.995
ZBSC0675	Lineage 2	0.005	0.995
ZBSC0678	Lineage 2	0.006	0.994
ZBSC0679	Lineage 2	0.012	0.988
ZBSC0681	Lineage 2	0.003	0.997
ZBSC0684	Lineage 2	0.004	0.996
ZBSC0685	Lineage 2	0.009	0.991
ZBSC0686	Lineage 2	0.008	0.992
ZBSC0688	Lineage 2	0.008	0.992
ZBSC0690	Lineage 2	0.004	0.996
ZBSC0691	Lineage 2	0.004	0.996
ZBSC0692	Lineage 2	0.004	0.996
ZBSC0693	Lineage 2	0.004	0.996
ZBSC0694	Lineage 2	0.004	0.996
ZBSC0695	Lineage 2	0.004	0.996
ZBSC0696	Lineage 2	0.003	0.997
ZBSC0697	Lineage 2	0.005	0.995
ZBSC0699	Lineage 2	0.005	0.995
ZBSC0703	Lineage 2	0.004	0.996

Annex 13. Structure bar plot of Bayesian assignments of the individuals to the respective cluster for K=3 and K=4. Vertical bars indicate individuals and the colours within each bar denote for the probability of membership of each specimen to a cluster.



Annex 14. Number of alleles sampled and gene diversity for each locus and population analysed.

Locus	Alleles sampled			Gene diversity	
	L1	L2	Total	L1	L2
Jac16	9	2	9	0,725	0,033
Jac18	9	18	20	0,831	0,913
Jac37	16	15	29	0,9	0,801
Jac02	17	18	21	0,859	0,886
Jac08	13	13	18	0,835	0,822
Jac23	13	11	19	0,757	0,78
Mean	12,83	12,83	19,33	0,82	0,71